MARCO BIAGI FOUNDATION / DEPARTMENT OF ECONOMICS MARCO BIAGI

PHD PROGRAMME IN "LABOUR, DEVELOPMENT AND INNOVATION"
XXXIV CYCLE

"People Management 4.0: competence and algorithms in the digital enterprise"

PHD COORDINATOR

Prof. ssa Tindara Addabbo

Department of Economics Marco Biagi/Marco Biagi Foundation

SUPERVISOR AND CO-SUPERVISOR

Prof. Tommaso Fabbri, Prof. Francesco Pattarin

Department of Economics Marco Biagi/Marco Biagi Foundation

THE CANDIDATE

Shahin Manafi Varkiani

smanafiv@unimore.it

Academic year 2020/2021

People Management 4.0: competence and algorithms in the digital enterprise

Shahin Manafi Varkiani

Abstract

The rise of the fourth industrial revolution has led to a radical transformation of business processes, organizational forms, quality of work and working conditions. Several studies have documented that the increase in the number of available information sources and the amount of data internal or external to organizations, together with the increasing availability and power of information storage and processing technologies, has made companies aware of how much automation and big data can represent a source of competitive advantage and a tool for the evolution of their business models. The technical change brought about by digitalization and other related processes affects the human component, on which the research project focuses.
This study aims to evaluate the effects of the digital transformation on the labor market and on the management of the workforce, investigating how digital transformation impacts the evolution of skills and exploring the areas and effects of the data-driven turn in human resources management. Therefore, this study aims at and intends to be useful to academics, for the results of empirical analyzes, to human resources professionals, for data-driven HRM scenarios and applications, to institutions and policy makers in understanding and managing changes in professions and skills in a context of rapid technological change.

L'avvento della quarta rivoluzione industriale ha portato ad una trasformazione radicale dei processi di business, delle forme organizzative, della qualità del lavoro e delle condizioni lavorative. Diversi studi hanno documentato che l'aumento del numero di fonti di informazione disponibili e dei dati interni ed esterni alle organizzazioni, insieme alla crescente disponibilità e potenza delle tecnologie di archiviazione e elaborazione delle informazioni, ha reso consapevoli le aziende di quanto l'automazione e i big data possano rappresentare una fonte di vantaggio competitivo e uno strumento di evoluzione dei propri modelli di business. Il cambiamento tecnico determinato dalla digitalizzazione e da altri processi connessi ad essa colpisce la componente umana, su cui si concentra il progetto di ricerca.
Lo studio si propone di valutare gli effetti della rivoluzione industriale sul mercato del lavoro e sulla gestione della forza lavoro, indagando come la trasformazione digitale impatta sull'evoluzione delle competenze ed esplorando gli ambiti e gli effetti della trasformazione della gestione delle risorse umane in senso data-driven.
Pertanto, questo studio si rivolge e intende essere utile ad accademici, per i risultati delle analisi empiriche, ai professionisti delle risorse umane, per gli scenari e le applicazioni di data-driven HRM, alle istituzioni e ai policy maker nella comprensione e gestione dei cambiamenti delle professioni e delle competenze in un contesto di rapido cambiamento tecnologico.

Ringraziamenti

Finire di scrivere la tesi è come finire di leggere un libro: non vedi l'ora di arrivare all'ultimo capitolo ma poi ti sembra di dover salutare un amico al quale ti eri affezionato che ti aveva tenuto compagnia per tanto tempo. Quello che per me conta non è solo il risultato, ma la passione e l'energia che sento nel percorso assaporando ogni singolo giorno che passa quando faccio qualcosa che mi piace.

Vorrei ringraziare innanzitutto i miei genitori, per avermi insegnato con il loro esempio tante cose positive e con il loro cattivo esempio tante altre da evitare.
Un ringraziamento particolare va ai miei amic*, che mi hanno sempre voluto bene per ciò che sono, con i miei pregi e difetti, mi hanno sostenuta nei momenti difficili e hanno condiviso con me la felicità nei momenti di gioia e qualche attimo di sana follia che rende la vita più saporita.
Un ringraziamento speciale ai Padri Francesco Cavallini e Iuri Sandrin e ai miei compagni di viaggio con i quali ho fatto un lungo pellegrinaggio in Terra Santa cinque anni fa. Ripercorrere l'esperienza del cammino nel deserto e della fatica del popolo d'Egitto verso la libertà mi ha aiutato a scavalcare il rimpianto che porta a tirarsi sempre indietro e a scoprire che solo perseverando si ottengono le cose importanti. Le testimonianze ascoltate dalle persone dei territori occupati della Palestina che pure nella sofferenza avevano una grande voglia di fare e fare del loro meglio mi hanno dato un grande insegnamento.
Ringrazio la mia compagnia d'improvvisazione teatrale, improvvisando ho imparato a pensare in modo creativo e che quello che conta non è aver la scaletta delle battute, ma saperle creare strada facendo.
Ringrazio il Professor Fantoni che è sempre stato per me una guida, mi ha insegnato a guardare le cose sotto una prospettiva differente dalla mia, e ha scommesso su questo percorso.
Ringrazio il mio tutor, il Prof. Fabbri per avermi guidata in questo percorso e per avermi insegnato molto, il mio co-tutor, il Prof. Pattarin, che è stato di grande aiuto per il mio terzo capitolo, nonché un ottimo compagno di lavoro, il Prof. Solinas per aver accolto le mie idee e avermi offerto un concreto supporto nella scrittura del primo capitolo, la Prof. ssa Scapolan per avermi dato qualche spunto nel redigere il secondo.
Ringrazio la Fondazione Marco Biagi e la Fondazione Giacomo Brodolini per avermi dato l'opportunità di intraprendere questo percorso.
Ringrazio i miei colleghi, con i quali si è istaurato da subito un rapporto di amicizia che ha reso il cammino molto più bello. Insieme abbiamo condiviso l'esperienza lavorativa e tanti bei momenti di svago.
Ringrazio i Gesuiti e il mio caro amico Shady, che mi hanno fatto capire che la vera via della felicità non passa attraverso il canale della forza o del divertimento, ma dalla sfida più difficile: la ricerca della vocazione.
Ringrazio infine me stessa, per essere riuscita a buttarmi in questa avventura, con grande entusiasmo, passione, apertura verso il nuovo, voglia di conoscere e di fare. Caratteristiche che mi hanno portata ad andare in Olanda partendo da zero e conoscere subito molte persone, buttarmi nelle situazioni, nei corsi, nei viaggi e nelle esperienze più svariate di svago e di lavoro.
Credo che quello che conta di più nella vita non è avere un lavoro che ti renda ricco o che ti porti a raggiungere un particolare obiettivo, ma fare ciò che senti nel profondo del tuo cuore essere qualcosa che ti appartiene, qualcosa che ha a che fare con te, con le tue passioni, attitudini e con i tuoi desideri più profondi e mi rendo conto che ho vissuto questo percorso come una vera e propria vocazione. Questo è stato l'unico e semplice ingrediente che mi ha aiutata a superare le difficoltà, a cercare sempre di dare il mio meglio, e a non mollare mai il cammino nonostante gli ostacoli.
Auguro a tutti di trovare la propria vocazione, personale e lavorativa.

Acknowledgments

Finish writing your thesis is like finishing reading a book: you can't wait to get to the last chapter but then you seem to have to say goodbye to a friend you had grown fond of who had kept you company for so long. What matters to me is not only the result, but the passion and energy that I feel in the path enjoying every single day that passes when I do something I like.

First of all, I would like to thank my parents for teaching me by their example many positive things and with their bad example many others to avoid.
Secondly, I would like to thank my friends, who have always loved me for who I am, with my strengths and weaknesses, have supported me in difficult times and have shared with me happiness in moments of joy and a few moments of healthy madness that makes life tastier.
A special thanks to Fathers Francesco Cavallini and Iuri Sandrin and to my travelling companions with whom I made a long pilgrimage to the Holy Land five years ago. Retracing the experience of the journey in the desert and the effort of the people of Egypt towards freedom has helped me to overcome the regret that leads to always pull back and find that only by persevering you get the important things. The testimonies heard by the people of the occupied territories of Palestine who, even in their suffering, had a great desire to do and do their best have given me a great teaching.
I am very grateful to improvisational theater company, I have learned to think creatively and that what matters is not to have the lineup, but to know how to create them along the way.
My gratitude is also extended to Professor Fantoni who has always been a guide for me, he taught me to look at things from a different perspective and he bet on this path.
I thank Prof. Fabbri for having guided me in this path and for having taught me a lot, Prof. Solinas for having accepted my ideas and for having offered me a concrete support in the writing of the first chapter, Prof. Scapolan for having helped me in writing the second one, and Prof. Pattarin, who was of great help for the methodological aspects of the third, as well as an excellent working companion.
I thank the Marco Biagi Foundation and the Giacomo Brodolini Foundation for giving me the opportunity to undertake this path.
I would like to thank my colleagues, with whom a friendly relationship has been established from the outset and which has made the journey much more beautiful. Together we shared the work experience and many nice moments of leisure.
I thank the Jesuits and my dear friend Shady, who made me understand that the true path of happiness does not pass through the channel of strength or fun, but from the most difficult challenge: the search for vocation.
Finally, I thank myself for being able to jump into this adventure, with great enthusiasm, passion, openness to the new, desire to know and do.
Characteristics that led me to go to Holland from scratch and get to know many people immediately, throw myself into situations, courses, travel and in the most varied experiences of leisure and work.
I believe that what matters most in life is not having a job that makes you rich or that leads you to reach a particular goal, but doing what you feel in the depths of your heart to be something that belongs to you, something that has to do with you, with your passions, attitudes and with your deepest desires and I realize that I have lived this path as a real vocation.
This was the one and only ingredient that helped me to overcome the difficulties, to always try to give my best, and never to give up despite the obstacles.
I wish everyone to find their vocation, personal and professional.

Contents

Introduction

The advent of the fourth industrial revolution has led to a radical digital transformation of companies that increasingly produces a change in business processes, which make possible to exploit the intelligence introduced by digital technologies in most company's activities. The term “Industry 4.0” is derived from an initiative launched by the German government for safeguarding the long-term competitiveness of the manufacturing industry (Kagermann *et al*., 2013; cited in Müller *et al.*, 2018). With the terms “Industry 4.0” and “Fourth Industrial Revolution”, institutions and researchers refer to the “transformation of production of goods and services resulting from the application of a new wave of technological innovations” (Caruso, 2018). According to Schwab (2016), founder and executive chairman of the World Economic Forum, “the Fourth Industrial Revolution is building on the Third, the digital revolution that has been occurring since the middle of the last century. It is characterized by a fusion of technologies that is blurring the lines between the physical, digital, and biological spheres”. In the same vein, Müller *et al.* (2018) argue that “this new approach leads to industrial value creation that is not only automated, mostly within single manufacturing plants, but also interconnected between objects, products, and humans, building on the concept of the Internet of Things”.

As Zazancoglu and Ozkan-Ozen (2018) observe: “all these changes […] are expected to alter the job profiles of employees in different kind of ways, and it is essential to focus that area as well”.

While some researchers argue that the digital revolution certainly has an extraordinary potential for human progress and offers several opportunities, others are concerned about the replacement of human work by machines. The optimists argue that the replacement of routine tasks, characterised by labor intensive and a medium level of cognitive effort required, by automation has made it possible to reduce the types of physically demanding, repetitive, dangerous and mentally monotonous tasks, allowing workers to devote themselves to performing tasks that require flexibility, creativity, problem-solving and communication skills (Author, 2015). Certainly, the rapid advances in technology have made it possible to automate many human tasks, while augment others and redefine the tasks of many more (MacCrory *et al.*, 2014; World Economic Forum, 2019). An example is the advent of machine learning, that has led to a large-scale worker dislocation: white collar workers are replaced by professionals working in areas such as speech recognition, pattern recognition and image classification (Cascio and Montealegre, 2016). The scope of the replacement of human labour by machines is limited, as there are many tasks that relate to the tacit and non-explicable component of knowledge, for which neither programmers nor anyone else can enunciate specific rules or procedures. This limitation is known as Polanyi's paradox, "*We know more than we can tell*" (Polanyi 1966; Autor 2015).
Nam (2019) found that technology usage and long-term job perceptions becomes critical to job insecurity perception, and categorized the consequences of job insecurity found in the literature: individual and immediate reaction (job attitudes such as job satisfaction and job involvement), organizational and immediate reaction (organizational attitudes such as organizational commitment and trust), individual and long-term reaction (physical and mental health), and organizational and long-term reaction (work-related behavior such as performance and turnover intention).
Moreover, as Bersin (2019) argues, technology is mutating job into new, unexpected forms, as “hybrid” jobs are emerging, which are types of professions in which skills sets that never used to be found in the same job are combined, such as marketing or statistical analysis. They are less likely to be automated than other roles.

As the biologist and American writer Zinn teaches, "You can't stop the waves, but you can learn to surf" (John Kabat-Zinn), is it necessary to learn to deal with the continuous changes in employment without each time producing trauma or to passing on to the community the costs that others have seriously generated. Foreseeing and anticipating the changes in industrial processes that affect work is not only possible, but essential for the future. The human component is indeed affected by the technical changes brought about by the digitalization and other related processes.

This study therefore aims to evaluate the effects of the digital transformation on the labor market and on the management of the workforce, investigating how digital transformation impacts the evolution of skills and exploring the areas and effects of the data-driven turn in human resources management.

This dissertation is composed of three chapters.

- In Chapter 1, starting from an analysis of the copious literature concerning the effects of the fourth industrial revolution on labour and from the analysis of classification and certification systems of competences, the main open questions are proposed for scholarly reflection. The essay suggests a new way of looking at the classifications of professions and skills in a context of rapid technological change.
- In Chapter 2, the state of the art of People/HR analytics is explored through a systematic literature review.
- Finally, Chapter 3 contains an empirical part that consists of a case study application of HR analytics to investigate and predict employee attrition in a large financial company.

References

Autor D. H. (2015), Why Are There Still So Many Jobs? The History and Future of Workplace Automation, Journal of Economic Perspectives, vol. 29, n. 3, pp. 3-30, doi:10.1257/jep.29.3.3.

Bersin, J. (2019). The Hybrid job Economy: How New Skills are Rewriting the DNA of the job Market. Burning Glass Technologies.

Caruso, L. (2018). Digital innovation and the fourth industrial revolution: epochal social changes?. *AI & Soc* 33**,** 379–392. https://doi.org/10.1007/s00146-017-0736-1

Cascio, W., & Montealegre, R. (2016). How Technology Is Changing Work and Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 349–375. https://doi.org/10.1146/annurev-orgpsych-041015-062352

Cipriani, A., Gramolati, A., Mari, G. (2018) Il lavoro 4.0. La quarta rivoluzione industriale e letrasformazioni delle attività lavorative. Firenze University Press, ISBN: 8864536485.

MacCrory, F., Westerman, G., AlHammadi, Y., & Brynjolfsson, E. (2014). Racing With and Against the Machine: Changes in Occupational Skill Composition in an Era of Rapid Technological Advance. *ICIS*.

Müller, J. M., Buliga, O., & Voigt, K.-I. (2018). Fortune favors the prepared: How SMEs approach business model innovations in Industry 4.0. *Technological Forecasting and Social Change*, *132*, 2–17. https://doi.org/10.1016/j.techfore.2017.12.019

Nam, T. (2019). Technology usage, expected job sustainability, and perceived job insecurity. *Technological Forecasting and Social Change*, *138*, 155–165. https://doi.org/10.1016/j.techfore.2018.08.017

Schwab, K. (2015) The fourth industrial revolution: what it means, how to respond. World Economic Forum. Available at: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/. Accessed 22 December 2021

World Economic Forum (2019). Towards a Reskilling Revolution. Industry-Led Action for the Future of Work. In collaboration with Boston Consulting Group. Centre of New Economy and Society Insight Report.

Kazancoglu, Y., Deniz Ozkan-Ozen, Y. (2018) "Analyzing Workforce 4.0 in the Fourth Industrial Revolution and proposing a road map from operations management perspective with fuzzy DEMATEL", Journal of Enterprise Information Management, https://doi.org/10.1108/EIM-01-2017-0015

Digitisation, local knowledge and job descriptions. Preliminary notes on Polanyi's paradox

Shahin Manafi, Giovanni Solinas
Dipartimento di economia Marco Biagi
Università di Modena e Reggio Emilia[1]

Sommario

Come è noto, il paradosso di Polanyi, recentemente riscoperto da David Autor, sottolinea che gli esseri umani conoscono molto più di quanto siano in grado di tradurre in codice. Nell'esercizio del lavoro, vi sono compiti nei quali prevalgono empatia, intuito, manualità, esperienza cumulata, percezioni, rapporti relazionali tra individui, ecc. Le professioni (o, più di frequente, aspetti particolari delle professioni) che hanno queste caratteristiche – che fanno, se si vuol dire altrimenti, ampio uso delle cosiddette "soft skills" – non sono riducibili a sequenze procedurali definite, non sono cioè codificabili. Ne discendono implicazioni immediate sugli effetti della Quarta Rivoluzione Industriale sui livelli e la struttura delle occupazioni: le professioni (o aspetti delle professioni) descritti non sono automatizzabili. Non si innesca, in altre parole, un processo di sostituzione di uomo con macchine. I possibili sentieri evolutivi e di sviluppo di singoli processi e di specifici sistemi produttivi, come evidenzia ancora Autor, dipendono in modo decisivo dal mix e dalle relazioni tra saperi codificati e non codificati che è proprio di ciascuno di essi. Questa conclusione determina esiti assai prossimi a molta della letteratura contemporanea sui sistemi regionali di innovazione e, più in generale, sullo sviluppo locale e regionale. Il paradosso di Polanyi, si sosterrà è un importante anello di congiunzione tra questi due filoni di letteratura, per molti versi assai lontani.

In questo saggio, in particolare, si guarda a questi due filoni di pensiero a partire da un particolare angolo prospettico: quello dei sistemi di classificazione delle professioni e delle competenze. Si argomenta che le due tassonomie sono soggette a tensioni profonde sia per effetto del progresso tecnico, sia per effetto delle specificità locali e, in particolare, dei saperi di luogo. Molti dei problemi che si osservano nei sistemi formativi a livello locale e nella stessa formulazione dei contratti di lavoro derivano proprio dall'interazione tra specificità di luogo e progresso tecnico. Tanto più in periodi storici di forte cambiamento della tecnologia e delle traiettorie di sviluppo locale. La tesi di fondo che viene proposta è che le esigenze di identificazione dei mestieri e delle competenze, di standard che ne consentano la loro trasferibilità tra i luoghi e, in ultima istanza, di tutela lavoro sono oggi in particolare difficoltà proprio per le straordinarie accelerazioni associate ai processi di digitalizzazione dell'economia e ai loro effetti specifici sulle singole economie.

Abstract

As is well known, Polanyi's paradox, recently rediscovered by David Autor, points out that human beings know much more than they are able to translate into code. In the exercise of work, there are tasks in which empathy, intuition, manual dexterity, accumulated experience, perceptions, relational relationships between individuals, etc. prevail. The professions (or, more frequently, particular aspects of the professions) that have these characteristics - which make, if it is said otherwise, extensive use of so-called 'soft skills' - cannot be reduced to defined procedural

[1] E-mail: smanafiv@unimore.it; giovanni.solinas@unimore.it.

sequences, i.e. they cannot be codified. There are immediate implications for the effects of the Fourth Industrial Revolution on the levels and structure of occupations: the occupations (or aspects of occupations) described are not automatable. In other words, a process of replacement of man by machines is not triggered. The possible evolutionary and development paths of individual processes and specific production systems, as Autor again points out, depend decisively on the mix and relationships between codified and non-codified knowledge that is specific to each of them. This conclusion leads to outcomes very close to much of the contemporary literature on regional innovation systems and, more generally, on local and regional development. Polanyi's paradox, it will be argued, is an important link between these two strands of literature, which are in many ways very far apart.

This essay, in particular, looks at these two strands of thought starting from a particular angle: that of the classification systems of occupations and skills. It is argued that the two taxonomies are subject to profound tensions as a result of both technical progress and local specificities and, in particular, local knowledge. Many of the problems that are observed in training systems at local level and in the formulation of employment contracts derive precisely from the interaction between local specificity and technical progress. All the more so in historical periods of strong change in technology and local development trajectories. The basic thesis that is proposed is that the need to identify trades and skills, standards that allow their transferability between places and, ultimately, job protection are in particular difficulty today due to the extraordinary acceleration associated with the processes of digitalisation of the economy and their specific effects on individual economies.

Keywords: Occupational classification, skills classification, tasks, Fourth Industrial Revolution, local knowledge.

1. Introduction

Questo saggio inizia con una favola: la favola semiseria di Primo, Seconda e ultimo alle prese con Ribelle e con l'Occhialuto. Iniziamo col presentare i personaggi e a dire qualcosa sugli interpreti. La riportiamo in italiano per poi proseguire, come nel resto del saggio, in inglese.

Personaggi e interpreti

Primo. È uno che ha il pallino della classificazione. In particolare, esamina, ordina, e mette in fila le occupazioni e le professioni. È un signore molto metodico e scrupoloso. Ha sempre da aggiungere, modificare e correggere. E pensa di non aver mai fatto abbastanza. È anche un signore buono che difende i lavoratori, accompagna i migranti. Un paladino dei diritti. Classificare, non solo lo fa perché deve, ma perché gli piace proprio, tenendo conto, quando può, dei cambiamenti e dei più minuti dettagli. Questa è la sua vera debolezza. Anche se ha una storia più antica, il primo dei Primo famosi si chiamava ILO. È molto amico di Sindacato, anche se è un'amicizia diventata un po' abitudinaria, un po' stanca. È sposato con Seconda (che in qualche caso lo segue e in qualche caso gli cammina davanti …).

Seconda. Seconda è l'alter ego di Primo. Ha con Primo un rapporto insolubile. Lo tiene a bada impedendogli di sognare troppo. Lo tiene insomma con i piedi per terra. Ma è anche lei molto pignola. Dice a primo dalla mattina alla sera: "Guarda che se quello lo chiami 'arrotino' deve saper affilare i coltelli. Se quell'altro lo chiami *plumber* deve saper riparare i rubinetti e forse anche un tubo, …" così infaticabilmente. Di tutto questo Seconda tiene memoria e, in questo, è tanto maniacale quanto Primo. Avete capito, Seconda si occupa di classificare le abilità e le competenze.

Non di rado diventa tutt'uno con Primo. Quando può e quando riesce dice sempre: "guarda che questo quel lavoro lo sa fare davvero"; e, se glielo lasciano fare, lo scrive. Quando lo scrive, dice, un po' pomposamente che sta certificando (… ognuno ha le sue manie e bisogna comprenderla). Non tutti, ovviamente, le credono.

Ribelle. Un altro personaggio della nostra storia è Ribelle. Ribelle è un fanatico di tecnologia: sa tutto. Riorganizza i lavori più svariati, fabbriche o uffici, non si tira mai indietro. Sa aggiustare i computer, ne inventa di nuovi, li sa far funzionare, sa cosa è l'intelligenza artificiale e conosce i big data. È un po' un maniaco, ma è molto operativo, sempre indaffarato, molto stimato. Spesso litiga con Sindacato (che è un po' rozzo e non sempre sa apprezzare il suo genio). È visto come la peste da Primo e Seconda. Scombussola loro la vita, fin da quando lavorava con "sarto", o meglio come lo chiamano gli anglofoni Taylor (che era a sua volta molto amico di uno che faceva macchine nere, tutte uguali ed era molto ricco). Per Primo e Seconda, Ribelle è un vero incubo. Se potessero lo ucciderebbero. Quando Ribelle lavora molto non riescono proprio a stargli dietro. Ribelle ogni tanto viene chiamato automazione e oggi, spesso, con qualche confusione, Industria 4.0. Ma lui lo considera un soprannome che gli ha dato una signora tedesca e si offende.

L'Occhialuto. Ribelle poi, e questo è il nostro quarto personaggio, si accompagna spesso a un altro personaggio che chiamano l'Occhialuto. Anche l'Occhialuto è una spina nel fianco di Primo e di Seconda. L'Occhialuto è un umanista, un intellettuale, sta sempre lì a dire a Primo e a Seconda "a mo' caro, ma *plumber* non è mica come idraulico! A casa mia le cose si fanno diversamente. Seconda non può mica raccontarmela così all'ingrosso. E poi… competenze, signora mia, ma di cosa parla? Ci sono quelle dure e ci sono quelle morbide. Un *plumber* deve essere anche uno che sa chiacchierare… Non chiamo uno qualsiasi a casa mia; deve essere pulito, simpatico, perbene (chi fa la cronaca di questi dialoghi è ovviamente emiliano). Occhialuto parla di cose difficili, molto di locale, di non codificabile, di luogo, di *soft*… lo capiscono in pochi. Dicono sia nato in Toscana e che si chiami Giacomo (NdA Becattini) … ma non è sicuro. Gli inglesi lo negano e pensano si chiami Michael (Polanyi). Ribelle nel profondo detesta l'Occhialuto. Ma entrambi, spesso insieme nonostante tutto, danno il tormento a Primo e a Seconda.

Ultimo. L'ultimo personaggio si chiama appunto Ultimo. È figlio di Primo e di Seconda. È un bambinone dalla personalità indefinita. Lui pensa, nel suo intimo, che nel futuro dovrà fare insieme il lavoro che fanno suo babbo e sua mamma. Li dovrà fare molto più in fretta (accidenti al Ribelle!!) E lo dovrà fare in modo molto più flessibile, tenendo conto dei saperi di luogo e di capacità impalpabili (accidenti all'Occhialuto!!).

Suo padre e sua madre sono molto in ansia. Gli ricordano sempre che bisogna unire, guardare al generale, tenere memoria; se no, la gente non capisce e qualcuno ci marcia. Non si stancano di dirgli che la cura con cui si fanno le cose va bene, ma non si può stare troppo dietro ai dettagli; si fa di tutto per essere aggiornati, certamente, ma non si può cambiare ogni minuto.

Gli interpreti. GS e SM di queste vicende molto complesse sono osservatori e narratori. Stanno cercando di capire come sarà tra qualche anno Ultimo. E come i suoi genitori sono molto preoccupati. Si chiedono se non sia opportuna una rottura con la tradizione e vadano lasciati perdere i canoni di ILO e degli Atlanti del lavoro sparsi nel mondo; se non ci sia alcun valore legale/diritto maturato da difendere: il mercato, con qualche correttivo, con più o meno vincoli normativi, nel bene o nel male, farà come gli pare. Si chiedono se di tutto il lavoro di Primo e di Seconda non debba rimanere qualcosa di generale e di molto più sobrio e essenziale. Si chiedono, infine, se possa essere risolutiva, anche in questo caso, la consulenza di Ribelle. Ma SM e GS, come si è detto, sono solo dei poveri narratori.

Fuor di metafora. Primo rappresenta, in senso lato, i sistemi di classificazione delle occupazioni. Seconda, che ne è inscindibilmente legata, rappresenta i sistemi di classificazione delle competenze. Ultimo (che fa la sua comparsa soltanto in sede di conclusioni) simboleggia i possibili sviluppi futuri

di Primo e Seconda. Ribelle è il progresso tecnico, nella specifica forma associata ai processi di trasformazione digitale dell'economia. L'Occhialuto ha in commedia diverse parti. La principale è quella che riguarda le specificità dello sviluppo locale. Entrambi – il Ribelle e l'Occhialuto– vengono chiamati in causa principalmente in relazione ai loro effetti sui sistemi di classificazione delle occupazione e delle competenze. Personaggi secondari (quali, ad esempio la certificazione delle competenze) fanno la loro comparsa soltanto in appendice.

This essay begins with a fairy tale: the semi-serious tale of Primo, Seconda and Ultimo grappling with Ribelle and the Occhialuto. We begin by introducing the characters and saying something about the performers. We report it in Italian and then continue, as in the rest of the essay, in English.

Characters and interpreters

Prime. He is a man with a bump for classification. In particular, he examines, sorts, and lines up occupations and professions. He is a very methodical and scrupulous man. He always has to add, modify and correct. And he thinks he has never done enough. He is also a good man who defends the workers, accompanies the migrants. A paladin of rights. He classifies, not only because he has to, but because he likes to, taking into account, when he can, changes and the most minute details. This is his real weakness. Although it has an older history, the first of the Famous Primes was called ILO. He is very friendly with Syndicate, although, it is a friendship that has become a bit habitual, a bit tired. He is married to Second (who sometimes follows him and sometimes walks in front of him...).

Second. Second is Primo's alter ego. She has an insoluble relationship with Primo. She keeps him at bay by preventing him from dreaming too much. In other words, she keeps him grounded. But she's also very fussy. She tells Primo from morning till night: "Pay attention, if you call that one 'a grinder', he must know how to sharpen knives. If you call that one a plumber, he must know how to fix taps and maybe even a pipe, ..." so tirelessly. Second keeps memory of all this and, in this, she is as obsessive as Primo. You understand, Seconda deals with classifying skills and competences. Not infrequently, he becomes one with Prime. When she can, and when she succeeds, she always says: "look, he really knows how to do that job"; and if they let her do it, she writes it. When she writes it, she says, a little pompously, that she is certifying (... everyone has their foibles and you have to understand her). Not everyone, of course, believes her.

Rebel. Another character in our story is Rebel. Rebel is a technology fanatic: he knows everything. He reorganises the most diverse jobs, factories or offices, he never backs down. He knows how to fix computers, he invents new ones, he knows how to make them work, he knows what artificial intelligence is and he knows about big data. He's a bit of a maniac, but he's very operational, very busy, very esteemed. He often quarrels with Syndicate (who is a bit crude and does not always appreciate his genius). He is seen as the plague by Prime and Second. He messes up their lives, ever since he used to work with, as the English-speakers call him, Taylor (who was in turn very friendly with a man who made black cars, all the same and was very rich). For Prime and Second, Rebel is a real nightmare. They would kill him if they could. When Rebel works hard they just can't keep up. Rebel is sometimes called automation and nowadays, often with some confusion, Industry 4.0. But he considers it a nickname given to him by a German lady and is offended.

The Bespectacled. Rebel then, and this is our fourth character, is often accompanied by another character they call the Bespectacled. The Bespectacled is also a thorn in the side of Prime and

Second. The Bespectacled is a humanist, an intellectual, and he is always there to tell Prime and Second, "my dear, but plumber is not the same as plumber! In my house things are done differently. Second can't tell me that story wholesale. And then... skills, my lady, what are you talking about? There are hard ones and there are soft ones. A plumber must also be someone who knows how to talk... I don't call just anyone to my house; he has to be clean, nice, decent (those who report on these dialogues are obviously from Emilia). Bespectacled talks about difficult things, a lot of local, of non-codifiable, of place, of soft... few understand him. They say he was born in Tuscany and that his name is Giacomo (Becattini) ... but it is not certain. The English deny it and think his name is Michael (Polanyi). Rebel at heart hates the Bespectacled. But both of them, often together despite everything, torment Prime and Second.

Last. The last character is called indeed Last. He is the son of Primo and Seconda. He is a big boy with an undefined personality. He thinks, deep inside himself, that in the future he will have to do together the work that his father and his mother do. He will have to do it much faster (damn the Rebel!!) and he will have to do it in a much more flexible way, taking into account local knowledge and impalpable skills (damn the Bespectacled!!).

His father and mother are very anxious. They are always reminding him that it is necessary to unite, look at the big picture, keep in mind; otherwise, people don't understand and someone will march in. They never tire of telling him that the care with which things are done is fine, but one can’t stay too hung up on details; everything is done to keep up to date, of course, but it cannnnot be changed every minute.

Interpreters. GS and SM are observers and narrators of these very complex events. They are trying to understand how Last will be in a few years time. And like his parents they are very worried. They wonder if a break with tradition is not opportune and if the canons of the ILO and the job Atlas scattered all over the world should be abandoned; if there is no legal value/right to be defended: the market, with some corrections, with more or less regulatory constraints, for better or for worse, will do as it likes. They wonder whether of all the work of Prime and Second should not remain something general and much more sober and essential. Finally, they wonder whether the advice of Rebel might be decisive, even in this case. But SM and GS, as has been said, are only poor storytellers.

Out of metaphor. Prime represents, in a broad sense, occupation classification systems. Second, which is inseparably linked to it, represents the skills classification systems. Last (which makes its appearance only in the conclusion) symbolises the possible future developments of Prime and Second. Rebel is technical progress, in the specific form associated with the processes of digital transformation of the economy. The Bespectacled has several parts in the play. The main one is about the specificities of local development. Both - the Rebel and the Bespectacled - are mainly called into question in relation to their effects on occupation and skills classification systems. Secondary characters (such as, for example, skills certification) only appear in the appendix.

Finally, very few considerations should be made about the authors' convictions and motivations for writing the essay.

The technical change brought about by the digitalization and other processes will have important effects both on the composition of employment and on the competences and ability requested from workers. The integration between economic systems, besides a role of its own related to trade

flows, contributes to disseminate the new technologies and acts as a multiplier, making the spread of new technologies even faster.

The aim of this essay is to understand how the fourth industrial revolution affects the labour market and how the institutions are managing this change through the adoption of occupational and skill classification systems to ensure that it does not adversely affect employability. To this end, the paper examines how the occupational and skill classification systems react to external shocks, with a particular focus on the fourth industrial revolution. Defining occupations/professions and competences becomes instrumental to understand which among them are decaying and are more likely to be substituted, which are changing and how, and which, finally, are emerging. In many cases, it is only a certain number of elements of a specific occupation that decline or emerge. The increasing interest that in recent years has been devoted to the occupational and skills classification systems comes from this framework. The ongoing trasformation's processes heighten the need to reflect on classification systems and to try to make them more effective. As always happens in these cases, a debate is taking place about the direction to follow. On the one hand, it might be thought that classification systems, in order to grasp the transformation processes, should be built around increasingly meticulous declaratories. On the other hand, it could be argued, in exactly the opposite direction, that a classification that is too detailed is always lagging behind the ongoing processes and, for the same reason, is not able to describe them, especially when change is very rapid.

The essay is structured as follows. Paragraph 2 summarises the main characteristics of the classification systems of occupations and skills, highlighting their historical genesis and function. In paragraph 3, 4 and 5 the effects of the revolution on labour markets are summarized with particular reference to the impact on classification systems. This is followed by an analysis of the relationship between specificity and knowledge of place and classification systems and the dynamics of local development (par. 6 and par. 7). The essay concludes with an examination of the outstanding and open questions (par. 8).

2. Skills and occupational classification systems

The systems for classifying professions and competences are as old as the systems of industrial relations, even the most archaic. The exchange of workforce against money has always required that the service be offered by those who are able to perform it and that the content of the service is defined. At least to some extent. Since the Twenties and from the seminal contribution of Coase (1937, 1988) it is, in fact, clear that they can only be incomplete contracts, precisely because a part of the contents of the labour services, which these contracts should regulate, cannot be codified. The job descriptions and the minute definition of the tasks that appear with Fordism and which last until the present day are, in fact, the attempt of the firm and of the firm holders, to remedy, with control over labour process to this irreducible incompleteness. On the contrary, for trade unions and international organizations, first and foremost the International Labor Office, meticulously defining tasks and designing validation systems of tasks, in the unequal exchange between capital and labor, aimed primarily at protecting the worker. All the more so in contexts of weak union and in labour markets with relevant migration. Workers' abilities and skills must be described, recognized and transferable. The evolution of goods markets changes the organization of work and, to some extent, undermines both visions.

Fragmentation and volatility of demand for products and increasing customization, in many production processes reduce standard procedures. The emphasis shifts from the intensity in the execution of an elementary task to the commitment and the participation of the worker. The Fourth

Industrial Revolution, – at least in jobs where it requires more abstract skills, relational and soft skills – goes in the same direction. The hypothesis put forward in these pages is that these transformations will have an increasing impact on the way in which the classification systems of occupations and skills will have to be re-constructed and re-thought.

Definition

An occupation can be defined as "a grouping of jobs involving similar tasks, which require a similar skills set" (ESCO, 2015, cited in Beblavy *et al.*, 2016). A job, on the other hand, "is bound to a specific work context and executed by one person" (ESCO, 2015, cited in Beblavy *et al.*, 2016).

As Beblavy *et al.* (2016, p.8) notes: "Occupations typically are presented in an occupational classification, in which they are grouped on the basis of similarity in terms of tasks, responsibilities, education and skill level".
In the definition of occupation and jobs listed above, the concept of task and skills are particularly relevant. Acemoglu and Author (2011, p. 2) define a task as "a unit of work activity that produces output (goods and services)" and a skill as a "worker's endowment of capabilities for performing various tasks". Skills are often proxied by occupations or derived from the occupational classifications (Beblavy *et al.*, 2016).
According to a definition provided by the International Labour Organisation (ILO), *occupational classification* is "a tool for organizing all jobs in an establishment, an industry or a country into clearly defined set of groups according to the tasks and duties undertaken in the jobs" (ILO, 2015), and therefore we can affirm that "classifications of occupations are a means for grouping jobs by their similarity" (Koucký *et al.*, 2012). According to Elias & McKnight (2001) "occupational classifications categorise the type of work that is performed in a job".
There is a link between occupational classifications and skills classification systems, since "standard skills classification systems function as a complementary rather than a primary source of occupational skills information" (Siekmann and Fowler, 2017). Skills classifications is more suitable than occupational classifications in reflect labour market transformation, the interdependencies between occupations and in reflecting new and emerging occupations (Siekmann and Fowler, 2017; Beblavy *et al.*, 2016). As the African Development Bank Group *et al.* (2018; cited in Hernandez, 2018, p. 5) point out, identify the skills and qualifications related to occupations "offers the possibility of performing workforce management for individuals based on the required skills and not just based on job titles or duties. This is particularly important in a changing labor market, where tasks performed by workers can shift with technology adoption".

History

Occupational classification systems have been attracting considerable interest since in the contemporary society there has been the need to understand occupational structure and mobility (Katz, 1972). In his analysis of the occupational classification systems' history, Katz identifies what were the ambitions that have led researchers to develop these systems: understanding "the distribution of the workforce between various kinds of employment at different points in the effect of changes, such as technological innovation, on that distribution" and facilitating "comparisons of the world of work, quickly, efficiently, and meaningfully. At the same time, they want to study mobility-the movement of individuals from occupation to occupation" (Katz, 1972, p. 64-65).

During the 19th century, the first national classification systems for occupation and industry were developed. They were initially very simple and consisted of a list of occupations not structured hierarchically and reflected social strata rather than tasks performed (A 't Mannetje, 2002). Recent

trends in international and inter-discipline comparisons and needs to matching jobs and workers led to the necessity to develop a uniform and standardized occupational language (A 't Mannetje, 2002 and National Research Council, 1980). For this purpose, during the International Conferences of Labour Staticians, in the 1923, the International Labour Organisation (ILO) discussed the need for an international standard classification of occupation (ISCO). As a result, in the 1949 concrete work to develop ISCO was initiated and in the 1957 the major, minor and unit group of the first ISCO were established (Hoffmann, 1999). The main purpose of this type of classification was to map national classifications into a common internationally comparable taxonomy, enabling in this way international comparisons and stimulating labour mobility (Beblavy *et al.*, 2016).

The systems in line with the innovations introduced at European level were initially promoted by the International Labour Office (ILO), which in 2008 launched the new classification of professions ISCO08 (International Standard Classification of Occupations, 2008).

Member States were invited to use the ISCO08 classification or a derived national classification as a basis for producing and disseminating statistical data on labour, but no provision has been made for a Community version of the international classification (Istat, 2013).

In Italy, in response to international recommendations, the National Statistical Institute has given a committee of experts the task of updating the previous taxonomy (CP2001).

Similarly to the new international taxonomy (ISCO08), CP2011 has not changed the underlying logic previously adopted: the criterion of competence, considered in its dual dimension of the level (skill level) and scope (skill specialization) of the skills required to perform properly the tasks associated with the profession (Istat, 2013).

ISCO has been adopted in Europe and other countries, and has developed specific variants to be applicable for each specific requirements: a European Union variant (ISCO-88(COM)), a commonwealth variant (ISCO-88(CIS)) and an Asian variant (ISCO-88(OCWM)) (A 't Mannetje, 2002 and Hoffmann, 1999). The ILO has actually developed the ISCO classification for two purposes (Van Leeuwen *et al.*, 2010): provide a systematic basis for presentation of occupational data relating to different countries in order to facilitate international comparisons and provide an international standard classification system which countries might use in developing their national occupational classifications.

The US developed occupational classification independently from Europe, because the ISCO was considered not flexible enough for the American context. They therefore created the Dictionary of Occupational Titles (DOT), which is nowadays replaced by O*NET, and, in the 1966, the Standard Occupational Classification (SOC). Based on these standard classifications, many countries have developed their own (A 't Mannetje, 2002; Moskowitz and Chief, 2017). The diversity and the variance in the occupational classification reflect a natural consequence of the differences in each country's history, law and regulations, labor market, technology, social and demographic change and economy (Moskowitz). For example, SOC lacks some occupations that ISCO has included because it is used in many countries in different stages of economic growth. Another representative case of this heterogeneity is the 1997 version of the Japan's Standard Occupational Classification that listed some detailed manufacturing occupations unique to its labor force, for example workers that make miso and soy sauce, tea, tofu, sake or kimonos and Japanese lantern and fan makers (Moskowitz). This variety between national occupational nomenclatures gives rise to occupational discordance, since the same individual might be classified as working in different occupations (and therefore of different skill levels) depending on where they are physically located (Parsons *et al.*, 2014).

For these reasons, it was very difficult to make a systematic international comparison, even if from 1841 onwards the standardization of occupational terminology became more marked. Many local terms were subsumed into broader categories, since the advantages of having a common system

for comparing and consolidating information from various sources outweigh the loss of precision of the multi-purpose classification (Martin, 1967), even if the search for one universal all-purpose coding of occupations has generally been abandoned by economic historians (London Electoral History). This happened because the grouping with which the available occupational information can be matched had to be flexible enough to allow for variations between one economy to another, as well as to permit a study of change over time (London Electoral History).

Role

Occupational classification systems are developed for a variety of purposes. Some of the main ones are: informs the job matching undertaken by employment agencies, provides career information for labour market entrants and yields guidance for the development of labour market policies (Elias & McKnight, 2001). Moreover, it produces intuitive sense and enables quick categorization of existing and potential new jobs whenever relevant (Van Vulpen, 2020).

Occupational classifications are therefore useful for (Hoffmann, 1999): legislators and public sector administrators, in support of the formulation and implementation of economic and social policies and to monitor progress with respect to their application; managers, for planning and deciding on personnel policies and monitoring working conditions; psychologists, who study the relationship between occupations and the personality and interests of workers; epidemiologists, in their study of work-related differences in morbidity and mortality; sociologists, for investigating of differences in life styles, behaviour and social positions and economists, that use occupation in the analysis of differences in the distribution of earnings and incomes over time and between groups, as well as in the analysis of imbalances of supply and demand in different labour markets.
The classification of skills has also received a lot of attention from researchers and policy-makers, since they realise that the traditional occupational classifications failed to reflect labour market transformation and interdependencies. They can be regarded as complementary to occupational classifications and can strengthen the link between the education sector and other sectors in the economy and stimulate mobility (Beblavy *et al.*, 2016).
In summary, the classification systems are developed by government agencies to carry out objectives such as[1]:
(i) To assist economists and data statisticians in their data collection efforts. Economists and statisticians use these systems when collecting census and other data such as data on worker mobility, technological change, and occupational employment statistics.
(ii) To analyze changes or patterns in the labor force and provide labor market information. Government agencies and organizations use these systems to understand changes in workforce demographics and other important labor market trends. This information is sometimes used to guide policy and develop systems for training, recruiting, and job matching. In addition, the information may be used to draw comparisons across work that, on the surface, may appear quite different.
(iii) To assist individuals in career exploration, career planning, and job seeking. It is important for job seekers, employment counselors, and employers to understand the requirements and descriptions of jobs and occupations so they may assist people in finding professions that match their skills and interests. Career guidance counselors use these systems to educate students or workers considering a career move or job transition. By matching job seekers' interests and level of knowledge and skill in job-related activities with those of various occupations, they can make an informed choice about a new career to pursue".
To achieve these goals, it is necessary to organize national mobility data to make comparisons between countries (Leeuwen, 2004). The differences, although we have seen them as a natural

consequence of the diversity of countries, do not help in translating from one country to another. A methodology that allows to "code" occupational information cross-nationally into a common classification scheme is needed to deal with the confusion due to the difference in occupational terminology across time and space, and between languages (Leeuwen, 2004) and the globalization of labour market has increased the demand for internationally comparable data for both statistical and administrative purposes (ILO, 2012). A method that allows to compare the coding of occupations is particularly useful to make the interoperability between systems less problematic.

Evolution of the systems, the upgrade path and the forces that drive change

Occupational classification and skills classification, that derive from the occupational classification, vary over time, as a result of technological and organizational change (Beblavy *et al.*, 2016; Katz, 1972).

The need to update and adapt the classification arises from the necessity to capture the changes in the labour market and in the occupational structure (Istat, 2013; Cattanei *et al.*, 2014). Another cause that led to the revision of these systems in order to develop taxonomies of highly aggregated major groups is the need to make reliable comparisons between countries (Cattanei *et al.*, 2014).
Some of the main forces driving these changes are: the innovation of production processes and their organization, the update in the qualification requirements required for the exercise of the professions and variation in the demand for goods and services (Istat, 2013). Other factors are: demographic trends, such as increased immigration, aging, and higher levels of education, business trends, shifts in consumer needs and preferences, laws, business practices and social developments, increasing scale of global competition and supply and demand of goods and services (Crosby, 2002; Elias & McKnight, 2001; Moskowitz & Chief, 2017). New occupations arises when employers need workers to do tasks that have never been done before (Beblavy *et al.*, 2016; Crosby, 2002). Initially, workers in existing occupations add tasks to jobs that already exist. But, as the difference between tasks grows, and become the primary jobs of enough workers, the "specialty" becomes an occupation of its own (Crosby, 2002; Beblavy *et al.*, 2016).
For example, the revisions that led the ILO to launch the new classification of professions ISCO08 level, to include emerging professions, in particular those affected by the impact of new IT and communication technologies, and to scale down the declining or severely reduced professional areas (Istat, 2013).
These transformations are quickened by the advent of the 'knowledge society', that brought new jobs or alteration of the cognitive contents and the tasks typically associated with some pre-existing professions (Cattani *et al.*, 2014).

As is easily understood, however, it is the transformations induced in the production processes and organisation brought about by the digitalisation of the economy that fully reveal the fragility of occupation and skills classification systems.

This is explored in the following section.

3. New technologies: jobs and employment

Our reasoning starts with an introduction on the rapid technological evolution that has characterized the last decade through the description of the two opposite points of view that have been created by the economists.

Mr Asimov's panglossian world: the optimistics

The mainstream vision is optimistic and claims that digitalization contains an extraordinary potential for humanity's progress and offers numerous opportunities. Among the many well-known descriptions, Asimov's is particularly meaninginful: *"In a properly automated and educated world, […], machines may prove to be the true humanizing influence. It may be that machines will do the work that makes life possible and that human beings will do all the other things that make life pleasant and worthwhile"* (Isaac Asimov, *Robot Visions*, 1990, p. 959). According to Asimov, in the future robots will allow humanity to become more human, because in the end they will conduct all those tasks that mankind does not like to do. Machines will therefore enable human beings to carry out activities according to their own nature which would prefer, for example, to express the arts rather that repair aqueducts. Everyone's lifestyle would thus become closer to everyone's needs and desires. In this vision, automation makes it possible to reduce the types of heavy activities on a physical level, repetitive, dangerous and monotonous on a mental level, allowing workers to devote themselves to performing tasks that require flexibility, creativity, problem solving and communication skills. This view is supported by many other scholars who, in the same vein, argue that the fourth industrial revolution will not collapse employment and, on the contrary, will have beneficial effects both in quantitative and qualitative terms.

In terms of quantity, data from several studies suggest that the corresponding growth in overall labour demand may offset the negative effects associated with the introduction of robots (Acemoglu and Restrepo, 2019, cited in Paba *et al.*, 2020). Paba *et al.* (2020) shown that there is a positive and statistically significant relationship between technology and employment, explained by the fact that the productivity increases expected from new technologies, together with the increased demand for more productive capital, can improve the efficiency of the industry, increase the demand for goods and services and contribute to the expansion of the economy also in sectors not directly interested in automation. Similarly, Cascio and Montealegre (2016) suggest that, according to the economic theory, the increase in productivity lead to a growth in the demand of new product and services, which, in turn, will create new jobs for displaced workers. In the same vein, the result of the studies of Petit (1995), Chennells and van Reenen (1999), Spiezia and Vivarelli (2002) (cited in Freddi, 2017) affirm that the effects of innovation on employment are positive, because innovative firms are more competitive and more productive, expand their markets and this allows them to grow more easily and to create new jobs opportunities.

A number of studies have shown that the digital revolution will lead to improvements also in the quality of work, replacing with machines all the heavy physical and mental jobs, in favour of activities more pleasant for mankind. Technology will allow an increase in creativity and autonomy.

According to Caruso (2018), technology is a supportive element for the release of routine work, in favour of creative, value-added activities and for the organization of work in a way that allow flexible working conditions that meet companies' requirements and personal needs of emoployees. In the same vein, the following examples show how certain professions benefit from technology. Thanks to support by artificial intelligence on diagnosing diseases, doctors, for instance, might spend less time on analysing symptoms and more time on ensuring a patient's well-being and individual needs. Similar screening tools will also be available for many professions and consultants. In many assembly lines, cobots assist operators by augmenting their capabilities in terms of effort, allowing them to manipulate parts that are hot, heavy, bulky, or too small for precision handling. In addition, with machines running around the clock, the workday of the operator is separate from that of the machine. The operator works in direct contact with the robot or in its immediate environment. With such proximity the operator can decide whether to interact or not with the machine. Operators

remain to control the machine and to deal with the most complex and non-standard operations. In different areas the same principles apply. Digital technologies allow data to be collected to analyze and interpret customer habits to develop a product or service that meets its needs and expectations. This strengthens the relationship with the customer that becomes continuous. Even after the purchase, in fact, it is possible to trace the data using IOT technologies to identify any critical issues and take timely actions to improve the customer experience.

Anti-panglossian world: the pessimistics

Other writers have argued that, in contrast to the Panglossians, the technological advancement can lead to the replacement of human work by machines, causing job losses and a dehumanization of human labour. The emergence of a huge increase in computing power, artificial intelligence and robotics has increased the possibility of replacing human work to levels never seen before. Machine learning techniques are expanding the range of replaceable tasks, they apply statistics and inductive reasoning where formal procedures are unknown. This has allowed the development of new applications in areas where human beings benefit: making predictions and making decisions about routine and non-mechanical activities. The processes of transferring work from workers to machines thus extend significantly also to activities traditionally considered difficult to automate by virtue of their cognitive and relational contents, such as banking management or retail.

It is certainly true that [...] "the workplaces in which the most innovative technologies are produced still depend to a large extent on human labor, while those in which traditional goods are manufactured are largely managed by robots" – as Moretti observes (p. 67); but it is not equally certain that the compensatory effects associated with technical progress are at least equal to the effects of job substitution. Certainly designing, implementing and maintaining these machines, computers and robots led to the emergence of a whole new industry. However, according to the International Labour Organization these activities offer significantly less employment opportunities than those lost in the process of automation (Ernst, Merola, Samaan, 2018).

These, and similar findings, (together with occupational trends in developed countries) generated the phenomenon labelled as "automation anxiety" (Akst 2013, Mokyr *et al.* 2015). The first cause of concern is the belief that technological progress will cause widespread substitution of human labour by mahines, which in turn could lead to technological unemployment.

From a quantitative point of view, therefore, the preoccupation is due to the belief that technological innovations lead to the destruction of a certain type and number of jobs (Brynjolfsson & McAfee 2014, Rotman 2013, cited in Cascio and Montealegre, 2016). A World Economic Forum's study predicts that 5 million jobs will be lost before 2020 (Caruso, 2018). Economists generally agree with the study of Autor and Dorn (2013), that concludes that the middle class is the most at risk of losing their job. A number of investigations have been been produced so far about the consequences of the application of the digital technologies on employment (Freddi, 2017), and they try to estimate the number of jobs that might be displaced by machines.

The ability of robots at substituting human labour has attracted the attention of many scholars. The reason is that the comporary robots are able to substitute not only the low-skilled repetitive tasks, but also more complex high-slill occupations (Freddi, 2017). This is exemplified in the work undertaken Ernst, Merola and Samaan (2018) that highlight the three main groups of tasks that have become the focus of Artificial Intelligence (AI).

In addition to the negative effects on quantity (job loss), some studies have shown that, under certain conditions, digitalization has a negative impact on the quality of work, such as totalitarian control, alienation and insecurity (Caruso, 2018). Fontana and Solinas (2021) analyzed working conditions in certain major companies in the province of Modena (Italy) and found that the technological development is having many negative effects on people's working lives. In particular, they suggested that health problems (musculo-skeletal disorders, work-related stress) are greater among digital workers, determined from the strong intensification of work and the standardisation of procedures and tasks, that leave unchanged or worsen the degrees of autonomy and control. In the same vein, Coovert & Thompson (2014b, cited in Cascio and Montealegre, 2016) notes that technology can be used to enable or to oppress people at work. They highlighted that the feeling of oppression occurs especially in cases where technology leads to a lack of autonomy, competence, and relatedness, causing stress, demotivation and counterproductive work behaviours.

The effects of industry 4.0 on employment: professions or tasks?

The following is a brief description of the studies present in the literature concerning the effects of industry 4.0 on employment. The literature is very broad, and, in these pages, we do not intend to provide a complete review. We will confine ourselves to considering the elements essential to our reasoning.
Frey and Osborne (2013) rang the bell for many other scholars. These authors estimate the probability of computerisation for 702 US occupations at one level. Their analysis is based on O*NET database. With the help of experts, they assessed a subset of professions on the basis of their level of automation, using discriminatory analysis methods to extend analysis to all profiles. Their unit of analysis are the occupational categories. They find that about 47% of total US employment is at risk, with a strong polarization of work (good and bad jobs). According to their study, transportation and logistics occupations, together with the bulk of office and administrative support are likely to be substituted by computers and other machines.

The results are consistent with other studies.
The same methodology has been applied to the study of the European situation (Bowles 2014), concluding that the proportion of workers who could be replaced by technological change is between 40% and 60%. According to this study, the countries that will be most affected are Romania, Portugal, Bulgaria and Greece. Similar results are obtained by Berger and Frey (2016) and by Chiacchio *et al.* (2018).

The results change significantly if the estimates on the effects on the employment of digitalization are studied not in relation to the professions, considered as a homogeneous whole for each category, but, instead, the tasks are considered. Arntz, Gregory and Zierahn (2016) estimate the job automatibility for 21 OECD countries through a task-based approach. They assume that it is not the entire occupation that is displaced by machines, but the specific job, depending on the tasks performed to complete it. So, jobs where workers perform a substantial share of automated tasks are more susceptible to automation than those with a higher share of non-automated tasks. This procedure makes it possible to differentiate the task structure within occupations and to focus on the specific task. This approach is therefore less restrictive than occupation-based approach, which is based on the assumption that task structures are the same across countries. They find that

applying a task-based approach results in a much lower risk of automation compared to the occupation-based approach. While Frey and Osborne find that 47% of US jobs are automatable, their corresponding figure is only 9%. The threat from technological advances is thus much less pronounced compared to the occupation-based approach by Frey and Osborne. Two examples might help to make the point. According to Frey and Osborne, people working in the occupation "Bookkeeping, Accounting, and Auditing Clerks" (SOC code: 43-3031) face an automation potential of 98%. However, only 24% of all employees in this occupation can perform their job with neither group work nor face-to-face interactions. Similarly, people working in the occupation "Retail Salesperson" (SOC code 41-2031) face an automation potential of 92%. Despite this, only 4% of retail salespersons perform their jobs with neither both group work nor face-to-face interactions.
Nedelkoska and Quintini (2018) based their study on the work done by Arntz, Gregory and Zierahn (2016) and extended their analysis to estimate the risk of automation to all the 32 countries that have participated in the PIAAC survey. They showed that about 14% of jobs are highly automatable and that another 32% of jobs have a risk of between 50% and 70% regarding the possibility of significant change in the way these jobs are carried out as a result of automation. This can happen, for example, in the case that a significant share of tasks, but not all, could be automated, changing the skill requirements for these jobs. They found that there is a large variance in automability across countries and that this variation is explained by the differences in the organization of job tasks within economic sectors.

Similarly, surveys that look at the effects of automation and robotization show a lower impact on employment than those based exclusively on the expected evolution of the professions (Acemoglu and Restrepo, 2017; Chiacchio *et al.*, 2018, Graetz and Michaels, 2018). These studies take into account the fact that, even if the presumed technological advances materialize, there are many other factors that influence the firms choose to automate.

Overall, these results suggest that the task-based approach appears to be more reasonable because it is more feasible to assume that the professions themselves do not lapse and that, instead, to decay and become obsolete are just some of the tasks that characterize a profession. Using information on task-usage at the individual level leads to significantly lower estimates of jobs "at risk", since workers in occupations with high automatibilities nevertheless often perform tasks which are hard to automate. Task-based analysis also justifies the difference in results between different countries, as people in different organisations, occupations and education group perform different tasks. This substantial difference is driven by the fact that even in occupations that Frey and Osborne considered to be in the high-risk category, workers at least to some extent also perform tasks that are difficult to automate.

The reflection on employment levels ends up being one with that on the professions. Reflecting on the professions, in turn, leads to focus on the content of the work: the tasks required to the worker and the ability of the worker to perform them, and therefore on the cognitive aspects, competences and skills. The study of tasks therefore plays a fundamental role in studying employment. The next paragraph then focuses on the role that tasks play in understanding and managing the effects of technical progress.

4. Tasks, knowledge and skills: Polanyi's paradox

For the investigation of the degree of complementarity and substitutability of man with machines, Autor, Levy e Murnane (2003) classify tasks as routine and non-routine, manual and cognitive (fig. 1). Routine and manual-cognitive dimension are two intertwining and orthogonal units of analysis through which it is possible to determine the degree of the human machine complementarity and substitutability. Routine tasks are those that can accomplished by following explicit rules.
The manual/cognitive dimension refers to the type of work, physical or intellectual.

	Routine tasks	**Non routine tasks**
Manual tasks	Routine – manual: picking or sorting, repetitive assembly.	Non routine – manual: janitorial services, truck driving.
Cognitive tasks	Routine – cognitive: record-keeping, calculation, repetitive customer service (e.g. bank teller).	Non routine – cognitive: forming/testing hypoteses, medical diagnosis, legal writing, persuading/selling, managing others.

Fig. 1: Classification of tasks performed by workers. Source: Elaboration from Autor, Levy and Murnane (2003).

Examples of routine activities include: the mathematical calculations involved in simple bookkeeping; the retrieving, sorting, and storing of structured information typical of clerical work; and the precise executing of a repetitive physical operation in an unchanging environment as in repetitive production tasks. Because core tasks of these occupations follow precise, well-understood procedures, they are increasingly codified in computer software and performed by machines. This force has led to a substantial decline in employment in clerical, administrative support, and to a lesser degree, in production and operative employment.
Tasks that can be described through an explicit and codeable procedure lend themselves to being good candidates for automation, since it becomes technologically possible and cost-effective to transfer them from a human operator to a machine, whether it is a manual or an intellectual task. This confirms the theory of labor market polarization, that is, the fall in demand for middle-wage jobs, while non-routine cognitive and manual roles are well defended. A greater mass of displaced workers is thus reallocated towards the tails of the occupational distribution (Acemoglu and Autor 2011 and 2012). This concept is explained by the Moravec's oddity, which is the discovery by artificial intelligence and robotics researchers that, contrary to traditional assumptions, high-level reasoning requires very little computation, but low-level sensorimotor skills require enormous computational resources. As Moravec writes, *"it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility"* (Moravec 1988, p. 15).
This happens because non-routine tasks are not understood sufficiently to be specified in a code because they refer to the tacit and unxplainable component of knowledge, for which neither computer programmers, nor anyone else can enunciate the explicit "rules" or procedures. This constraint is known as Polanyi's paradox, *"We can know more than we can tell"* (Polanyi 1966, p. 4; also quoted in Autor 2014 and 2015). According to Polanyi, people are not often aware of the knowledge they possess and use when they perform certain activities. The transfer of this type of knowledge generally requires personal contact and trust, therefore it can become assimilated by actors working in close proximity to each other.

The term *tacit knowledge* is attributed to Michael Polanyi in 1958 in *Personal Knowledge*. In his later classical work *The Tacit Dimension*, he states that "we can know more than we can tell" (1966, p.4), that lies at the heart of his distinction between tacit and explicit or codified knowledge (Gertler,

2003). “He emphasizes that we can often know how to do things without either knowing or being able to articulate to others why what we do works” (Grant, 2007). The implications of Polanyi's work will be discussed at a more detailed level in the following pages. At this point, it is sufficient to focus on the aspects most immediately related to the digitalisation of the economy, as highlighted by Moravec and more recently developed by Autor, Dorn, Levy, Murnane, and others.

When we break an egg over the edge of a mixing bowl, identify a distinct species of birds based on a fleeting glimpse, write a persuasive paragraph, or develop a hypothesis to explain a poorly understood phenomenon, we are engaging in tasks that we only tacitly understand how to perform. Following Polanyi’s observation, the tasks that have proved most vexing to automate are those demanding flexibility, judgment, and common sense—skills that we understand only tacitly. Polanyi’s paradox also provides an explanation to the Moravec oddity: high-level reasoning uses a set of formal logical tools that were developed specifically to address formal problems: for example, counting, mathematics, logical deduction, and encoding quantitative relationships. In contrast, sensorimotor skills, physical flexibility, common sense, judgment, intuition, creativity, and spoken language are capabilities that the human species evolved, rather than developed. Formalizing these skills requires reverse-engineering a set of activities that we normally accomplish using only tacit understanding (Autor, 2015).

In line with Polanyi’s observation, Levy and Murnane suggest that computers cannot completely replace humans, because they are able to follow instructions but not to recognize patterns. They underline that *“as the driver makes his left turn against traffic, he confronts a wall of images and sounds generated by oncoming cars, traffic lights, store fronts, billboards, trees, and a traffic policeman. Using knowledge, he must estimate the size and position of each of these objects and the likelihood that they pose a hazard […] the truck driver [has] the schema to recognize what [he is] confronting. But articulating this knowledge and embedding it in software for all but highly structured situations are at present enormously difficult task”* (Levy e Murnane, 2004, p. 28, see also Freddi, 2017).
The same authors also suggest that computers cannot replace humans in complex communications: *“Conversation critical to effective teaching, managing, selling, and many other occupations require the transfer and interpretation of a broad range of information. In these cases, the possibility of exchanging the information with a computer, rather than another human, is a long way off”* (Levy e Murnane, 2004, p. 29).
Autor, Levy and Murnane (2003) characterize non-routine activities that cannot be automated in two main groups. The first are those that require problem solving skills, intuition, creativity and persuasion, which are defined as "abstract", typical of professional, technical and managerial figures. The latter include the ability to adapt to situations, to fluently communicate languages, and interpersonal relationships, typical of the catering sector, cleaning services, health care, protection and safety trades.
Autor *et al.* (2003) argue that there is a substitutive relationship between human capital and technology in the presence of both manual and cognitive routine tasks. There will be complementarity for less predictable tasks requiring analytical and social skill and decision-making but at the same time benefit from the support offered by the technologies in the creation and management of the contents and informations in digital format. Finally, automation of routine tasks neither directly substitutes for nor complements the core jobs tasks of low education occupations—service occupations in particular—that rely heavily on “manual” tasks such as physical dexterity and flexible interpersonal communication (Autor e Dorn, 2013).

Service occupations at issue are the least educated and lowest paid categories of employment that involve assisting or caring for others, for example: food service workers, security guards, janitors and gardeners, cleaners, home health aides, child care workers, hairdressers and beauticians, and recreation occupations. The share of US labor hours in service occupations grew by 30 percent between 1980 and 2005 after having been flat or declining in the three prior decades. This rapid growth stands in contrast to declining employment in all similarly low-educated occupation groups, which include production and craft occupations, operative and assembler occupations, and transportation, construction, mechanical, mining, and farm occupations. The increase was even steeper among noncollege workers, by which we mean those with no more than a high school education. With the decline in manufacturing, the services sector has taken on the role of generating jobs. Business services, transport and distribution offered new, tailor-made jobs for the best educated and qualified workers.

Autor and Dorn hypothesize that occupational polarization is driven by the interaction between two forces: consumer preferences which favor variety over specialization; and *non-neutral* technological progress, which greatly reduces the cost of accomplishing routine, codifiable job tasks but has a comparatively minor impact on the cost of performing in-person service tasks. According to the authors, non-neutral technological progress concentrated in goods production (non-service occupation activities) has the potential to raise aggregate demand for service outputs and ultimately increase employment and wages in service occupations. “Winning professions” are those performed by highly-educated workers, which use "abstract" skills such as creativity, problem-solving, and coordination (Autor and Dorn, 2013).[2]

Autor and Dorn’s view, technological progress takes the form of an ongoing decline in the cost of digitalizing routine tasks, which can be performed both by computer capital and low-skill (“noncollege”) workers in the production of goods (Autor and Dorn, 2013). Automation of routine tasks neither directly substitutes for nor complements the core jobs tasks of low-education jobs that rely heavily on “manual” tasks such as physical dexterity and flexible interpersonal communication. What is crucial, once again, is the nature of the activities carried out by the worker, who uses tacit, contextual and embodied knowledge in their ordinary activities.[3]

[2] An interesting case of failure, at present, is that of financial advisors. Several experiments have been carried out to replace financial advisors with artificial intelligence applications.

Often the setting is holistic: the AI advisor takes care of the customer, advises him not only when and how to invest, but to buy or sell the house, to have children or not and so on. Important institutions that have made significant investements in these tools are reconsidering them. At most, these tools are considered an aid, but beware of full customer custody. Perhaps two considerations are sufficient to understand the reasons for this reverse. The first one agrees with many financial advisors and can be summarized approximately as follows: for every hour spent with a customer, ten minutes are devoted to technical advice, the rest of the time is spent talking about the facts of life. The second explains the previous one and can be expressed according to Sigmund Freud: “Money matters are treated by civilized people in the same way as sexual matters--with the same inconsistency, prudishness, and hypocrisy” (Freud, 1913, cit. in Ferrari et. al, 2019, p. 62). The management of an estate is an eminently private matter which is not discussed with the first come, especially not with a machine. In other words, the profession of financial advisor, as it is today in the western world, must be founded on an extraordinary relationship of trust. It requires extraordinarily high relational skills: one of those soft skills that is particularly difficult to replace with a machine and with the algorithms that govern the machine. Similar considerations can be applied to many other professions (doctors, psychotherapists, lawyers, etc.).

[3] Taking from a conveyor belt and packing a tile is something that anyone can do. And that is certainly automatable. But very few are able to distinguish a perfect tile from a failed one and the automated control systems known to date continue to fail.

Summing up. technological transformation processes are seriously questioning established paradigms and historical practices, not least the system of industrial relations. The interconnection and integration between digitization (the set of devices and sensors capable of transmitting and processing a huge mass of data at a speed until now unthinkable) and automation (availability of robots capable of replacing men's work with greater speed and productivity) have revolutionised production processes, enabling faster and more flexible production and have led to greater customisation of production. In a dynamic and complex context, government and the public sector play a fundamental role in helping people to manage transformation under way and to ensure them better lives and better jobs. Anticipating changes in production processes that affect work is not only possible, but essential for the future (Cipriani *et al.*, 2018). As the american biologist and writer John Kabat-Zinn teaches, *"You can't stop the waves, but you can learn to surf"* (John Kabat-Zinn, 2005).

5. Work digitization and new job profiles

In addition to and beyond the expected effects on employment levels and on the quality and conditions of work discussed in the previous pages, the digitisation of work has a direct and profound impact on organisational models and on the worker's own roles in the work process. This, as will be exemplified shortly, has an obvious impact on occupational declarations and job descriptors. The faster and more pervasive the process of transformation (as in the case of digitisation), the more evident the fragility of the classification systems of employment and skills in use becomes. That both systems suffer a loss of explanatory power does not depend solely on the deterministic link with which, as noted, they are constructed. There are deeper mechanisms at work, which apply to all the great epochs of transformation, and which are worth reflecting on.

In fact, when it is claimed that a demand for new knowledge and skills is emerging, it is in fact pointing to the need for new professional figures capable of filling new roles, in many cases not yet well defined. Similarly, the decline in demand for certain skills coincides with that for professional figures and roles. This very rarely means that a "trade" disappears. Much more frequently, it means that work content, ways of working, roles and professional figures change. Skills, in other words, exist in the meantime insofar as they are embedded in people, in what men or women, young or old, educated or illiterate, do or are capable of doing (Campagna *et al.*, 2017 and 2019; Pero, 2019). Forgetting this, as will be discussed extensively in the conclusion, means not being able to design adequate training pathways. And it is of course in capturing the ability to do something that the descriptors need to work.

The key point is therefore what we can reasonably assume today about what will be the evolution of functions and roles following digitisation. Certainly, new roles of a technical-specialist nature will appear (and are already appearing), requiring flexibility, the ability to adapt to different situations and environments, the ability to use different technologies, and to mix 'high' knowledge deriving from different disciplines. At the other extreme, many jobs will continue to be standardised and repetitive with little power of control on the part of workers (Fontana and Solinas, 2020; Garibaldo and Rinaldini, 2021).

For most people, for the so-called 'average' worker, roles will change to a very significant extent. Professions and trades will not disappear, as suggested by much of the literature emphasising the replacement of work by machines and algorithms, but what is done today will also be done

tomorrow, but in a different way. Some roles will simply be enriched, in the sense that digital skills will be added to traditional roles. For example, teaching in the classroom is one thing, teaching online is not exactly the same. It requires additional skills (e.g. knowing how to use one or more platforms, knowing how to communicate with a relationship mediated by a computer and a camera). For other roles the change will be more radical. The recurring example is that of the maintenance technician: with predictive maintenance, all maintenance is done remotely, computerised and digitised. And the in-person control and intervention on the machine will also change to some extent. So, this aspect changes drastically. And to a certain extent, the work on the machine, which uses a lot of digital instrumentation, also changes. The same thing is true for the car mechanic: in order to push the engine harder, to 'tweak', as they used to say, the cylinders and the bore are no longer affected, but the electronic control unit is. Similar considerations apply to the construction worker or the farmer. The spread of interdisciplinary roles is also appearing at lower levels of the employment scale. Even operating an advanced automation machine on an assembly line requires bringing together people who know how to do different things: operators, technicians, plant suppliers, engineers, artificial intelligence experts and those who will then operate the machine. Even those who clean in a factory or hospital must have new and different skills. For example, he/she has to know how to read on the tablet what kind of environment he is cleaning, what kind of sanitisation he/she has to do, what products he/she has to use, what precautions he/she has to follow, etc. and how he/she has to certify what he/she has done (Pero, 2019).

Today, many of the main technologies that are driving digitisation are well known (e.g. the so-called enabling technologies) but have very differentiated and evolving applications at the enterprise level. This also contributes to making them transversal and common to most occupations and at the most diverse levels of the occupational scale (Magnaghi, 2020).

The effects on occupational and skills classification systems are for these reasons very far-reaching.

Trade union bargaining and new professional roles: the metalworkers' contract

The perception of the processes of change (and their extent) is clearly evidenced by the fact that some of the elements discussed are also starting to emerge clearly in employment contracts. An emblematic example in Italy is that of the national contract for metalworkers (Negri and Pigni, 2015; STTR Tax & Labour 2021). The contract, described in more detail in Appendix 4, has two elements that, in relation to what is discussed here, are of great relevance.

The first is the transition from task to role. The historical system of occupational level descriptors, the declarations that finely describe the tasks that workers classified at a particular level can/must do, are abandoned in favour of much broader descriptors. There is a shift from the description of a set of fixed and stable tasks to an attempt to identify capabilities in a flexible and dynamic form. The second is to define the role as a mix of different competences[4]. Roles, in other words, are defined in a modular way. For each role, in fact, a fixed and a variable part is defined, which changes according to the company context. In this sense, the role can be defined by differentiated tasks (and capabilities). This construction is fundamental to recognise the new professionalism: modularity, understood in this way, puts the definition of professional profiles on a new basis. The abandonment of the job also marks the abandonment of a model rooted in a Fordist-Taylorist type of work organisation (and therefore also a model of industrial relations). This is clearly an epoch-

[4] To date, six have been identified, but it is highly likely that, due to the changes taking place, their number will increase.

making change. There is a clear move away from the logic that has guided the way in which classifications of professions and skills have been constructed. And it is not surprising that this proposal has arisen in a set of industries in which the mechanical industry, with the spread of mechatronics and the pressure to innovate production processes, is very strong.

6. Local knowledge and local productive systems

There is a further issue on which we believe that reflection should be opened, and which is closely connected to what has just been addressed. You can put it this way: a Sardinian, Texan and Modena lathe operator are all in the same professional declaratory. But they are not the same. The reason is the one mentioned in the opening. Every cognitive act, reminds us of Saul Meghnagi, is a response to a set of specific circumstances. "Learning can […] be considered as a socially situated activity within the framework of a particular reality of work of residence, of life" (Meghnagi 2005, p. 10, our translation). A copious literature on local development underlines how local knowledge, part of which has a tacit component, is a fundamental element of local competitiveness (for everyone Becattini, Brusco and Rullani). It is our opinion that this element continues to be relevant even if, as a result of new technologies, physical places are no longer the only container of organizational processes.

The role of tacit knowledge on local competitiveness

Each local system achieves an integration of explicit (codified) and tacit (contextual) knowledge (Becattini and Rullani, 1993) and successful economic systems are those in which the two spheres of knowledge interact continuously with each other, one feeding the other (Brusco, 1994).

In this paragraph we focus on tacit knowledge, for its relevance in the thesis we are supporting.
Again, the starting point of a very wide literature that includes the main scholars of local development in Italy and elsewhere, a relevant part of economic geographers and regional economists is the work of Michael Polanyi.
In explaining why “we know more than we can tell”, there are at least two distinct ideas connected to the statement (Gertler, 2003).
First is the issue of awareness or consciousness (Gertler, 2003). According to Polanyi, people are not often aware of the knowledge they possess and use when they perform certain activities. For example, when we break an egg over the edge of a mixing bowl, identify a distinct species of birds based on a fleeting glimpse, write a persuasive paragraph, or develop a hypothesis to explain a poorly understood phenomenon, we are engaging in tasks that we only tacitly understand how to perform and “the aim of a skillful performance is achieved by the observance of a set of rules which are not known as such to the person following them” (Polanyi, 1958, p. 51).
The second idea pertains to “communication difficulties and the inadequacies of language in expressing certain forms of knowledge and explanations, even when one has achieved full self-awareness” (Gertler, 2003, p.77).
Polanyi (1958) identifies the relationship between the master and the apprentice as the best way to transmit the hidden rules: “An art which cannot be specified in detail cannot be transmitted by prescription, since no prescription for it exists. It can be passed on only by example from master to apprentice. […] By watching the master and emulating his efforts in the presence of his example, the apprentice unconsciously picks up the rules of the art, including those which are not explicitly known to the master himself” (Polanyi, 1958, p. 55).

The transmission of tacit forms of knowledge takes places at best between people who have common characteristics such as speaking the same language, sharing the same rules and conventions, getting to know each other through past experience of informal collaborations or interactions. For these reasons, tacit knowledge must be learned by demonstration, imitation, performance and shared experience and therefore spatial proximity is the key to effective production and transmission/sharing of this type of knowledge (Gertler, 2003).
Tacit knowledge is therefore particularly present in local entrepreneurship, since it is owned and accumulated over time in a unique cultural context and is transmitted and exchanged through informal channels, through proximity and interactions between people, made possible by geographical continuity (Belussi, 1999; Ferrari, 2015; Hamel 1991). This type of knowledge can therefore be socialized directly only through long and costly processes of sharing the context and experiences (Becattini and Rullani, 1993), the values, language and culture (Maskell and Malmberg, 1999; Gertler, 2003). Tacit knowledge is embodied in individuals and in the collective learning of organizations and is only spread among citizens/workers in this specific context, remaining quite inaccessible to people from outside. It is territorially incorporated and characterizes the productive culture of every local production system that grows cumulatively over time (Belussi, 1999). It is also central in the process of learning-through-interacting that characterizes the geography of innovative activity and helps to reinforce the local over the global (Gertler, 2003).
Rullani (1994) clearly summarizes the difference between the two types of knowledge and their relationship with the socio-cultural context. He observes that contextual knowledge qualifies both as a specific expression of the socio-cultural environment in which the company is stablished, and as knowledge directly related to the contribution of human work to the quality of production processes. Codified knowledge, on the contrary, qualifies as an anonymous set of notions transferable even in different socio-cultural contexts through shared specialized languages.
The concept of tacit knowledge has received much attention in recent times because is related to the nature of modern competition (Gertler, 2003), especially in terms of the quality and reliability of the products (Brusco, 1994). The idea is that, in a world where everyone has easy access to explicit/codified knowledge, the creation of unique capabilities and products depends on the production and use of tacit, and spatially much less mobile, forms of knowledge (Maskell and Malmberg, 1999; Gertler, 2003).
The studies by Barney, Camuffo and Rumelt help us to understand why tacit knowledge is a source of competitive advantage for the competitiveness of places: the resources inside the enterprise constitute a competitive advantage if they are: rare, a source of value, durable, inimitable (Barney, 1991). The level of complexity of tacit knowledge plays a fundamental role in inhibiting or preventing imitation and more generally the transfer of knowledge (Camuffo, 2011; Rumelt, 1991). For these reasons, according to Becattini and Rullani (1993), the Local System is a place of accumulation of productive experiences and of new knowledge, and these are precisely the critical resources of the development of contemporary industrial capitalism.

The crisis of Fordism and the importance of contexts and their specificities

The importance of contexts and their specificities grew as the Fordist solutions, which in the past decreed the worldwide success of a few technological and organizational standards, lost force.

Today, in fact, in many cases the competitive advantage is achieved thanks to the diversity and adaptability in the production of goods, which is done better by small local companies, compared to large Fordists organizations (Becattini and Rullani, 1993). For this reason, the authors argue that the importance of contexts and their specificities grew as Fordist solutions lost their importance.

They claim that the variety of places of production play an essential role in the generation of competitive advantages, in the production of certain goods, compared to the same large Fordists organizations. This happens because large global organisations have reduced their ability to understand and use local specificities. Even if they find channels of connection with the environments in which they operate, the global enterprises establish, as a rule, with the places in which they operate and with the relative populations, less intimate and stable ties than they do, for example, the small and medium enterprises of a district (Becattini and Rullani, 1993). Local systems have very strong specificities which need to be taken into account in order to carry out the necessary mediation and integration between codified knowledge and tacit knowledge, which is necessary to achieve high levels of competitiveness and innovation capacity, and this applies to both small and large enterprises (Brusco, 1994). For this reason, also multinationals must take account of national (socio-cultural and institutional) contexts in order to maintain their competitive capacity, and are divided into smaller and more differentiated units, which can nest and grow within the different and peculiar local systems, to derive a more conscious and rich production style (Brusco, 1994).

The static and Fordist nature of the systems of occupational and competence classification systems comes into crisis when it has to respond to the complexity of today's world: *"They are the crisis of Fordism, the valorization of human resources, the organizational and functional flexibility, the transformation of working roles (most widespread and integrated), the destandardization of tasks, the IT and telematics revolution, the tertiary sector, globalization, innovation and all that we can trace back to the transition to post-fordism to shift the focus from work to the worker, from qualifying the role to the qualification of the one who is called to play that role"* (Lodigiani, 2011, p. 7).
A strong argument is being supporting: the logic on the basis of which job classifications were constructed and the same mechanisms for certifying skills were based, in essence, on the Fordist-Taylorist type organization of labour. And they still do reasonably well for those workers - many who have standardized tasks. The digitalisation of the economy and the automation of processes at the lower levels of the employment scale continue to generate fragmented, monotonous and repetitive jobs (Fontana, Paba e Solinas, 2019). In this context, work can be divided into standardized tasks (Fretwell *et al.*, 2001) and competencies are considered as a minimum unit of the breakdown of job tasks, associated with a specific occupation and role within an organisation (Lodigiani, 2011). However, as Stinchcombe (1990) argues, these types of workers are usually called “semiskilled” and their task skills do not form a large repertoire.

But when this does not happen, for jobs that require relational and transversal skills, attitude and habit to problem solving and so on, the acceleration of technical progress makes even more evident the fragility of occupational and skills classification and certification systems. This perspective ends up considering improperly identical or even indifferent the acquisition paths and ignoring the real contents of knowledge connected with practical skills (Meghnagi 2006, p. 282; cited in Lodigiani, 2011), which are generally acquired through learning by doing, learning by using, learning by trying, learning through alliance, and learning specially designed on-the-job training (Jin and Stough, 1997). What matters is not so much the formal level of qualifications achieved to promote the employability of the individual, but the complex of his cognitive, social and relational skills, which allow him to place and make spendable the skills possessed (Lodigiani, 2011). The notion of competence is described as an integrated set of knowledge, skills and attitudes and therefore becomes more complex (Pellerey, 2001, cited in Lodigiani, 2011) and tacit knowledge and skills become relevant in the professional context (Lodigiani, 2011).

Even in the case of job that requires routine tasks, defining the required skills is not as simple and immediate as it seems. The competences in turn incorporate other elements that make their definition complex. As Stinchcombe (1990) points out, competences usually have many components, as for example speed and accuracy within the routines and ability to switch among routines depending on the situation[5].

The solicitations arising from the individual production realities at the local level therefore play a role of tension with respect to the pressure for generalization brought about by the classification, making it even more difficult to define a coding standard in step with change and with geographical differences. As Lodigiani observes: "One important initial consideration is the tension between the standardization and formalization of recognition systems on one hand and the flexibility and capacity to adjust to each single recognition case, on the other" (2017, p. 136).

If local knowledge is such a powerful idiosyncratic element that it affects the same potential for growth of a productive system, this also has significant implications for the ways in which work can be described and represented. Perhaps the most significant example is that of the regional vocational training and skills certification systems. This is discussed in the following paragraph.

7. Vocational training, local knowledge and classification systems

In order to complete the reasoning proposed in the previous pages, it is useful to quickly review some aspects of the vocational training system. Italy is taken as a paradigmatic example of deep territorial differences that continue to persist, even in centrally defined systems.

The history of vocational training

In Italy, the history of vocational training is complex and rich, and it goes through different periods characterized by debates and reforms.
The phase that characterizes today's system began in 1986 with the Conference of Confindustria held in Mantua and brings to general attention the problem of Vocational Training, previously considered the subject of debate in a niche frequented by operators and experts.
This phase gives rise to regulatory production at national level but above all at regional level. From this moment on, the regional system, which had previously dealt almost exclusively with the first post-compulsory qualification, is increasingly concerned with school-educated individuals (high school and university graduates), in a strategic logic aimed at productive development (Ghergo, 2011).
Starting from this period, vocational training, which had already taken on different connotations from one region to another since the 1970s, has been accelerated and accentuated by the differentiation in terms of planning of activity plans, policy on the activities to be defended, the financial aspects and the denominations and duration of training courses.
The differentiation takes place in a double way: between educational offers and between regional systems. It is not only appropriate but necessary, as the activities and the professional specializations are very different from one region to another (Ghergo, 2011). To date, the education and training system involves the State and the Regions. "The former has exclusive legislative

[5] It has been argued that for these reasons, strict job classification contributed to the decline of the U.S. manufacturing productivity. This happened because some assembly plants detailed too much the work tasks, and therefore the limited flexibility did not allow workers to perform tasks outside contractual declarations. (Degen, 2011).]

competence with regard to the general rules on education and the determination of the essential levels of benefits to be guaranteed throughout the national territory, whereas the Regions have exclusive competence in vocational education and training. In more detail, the actors involved in the governance of the education and training system in Italy are: the Ministry of Education, University and Research (MIUR)[6] with tasks of general definition of the principles and essential levels of the education system; the Ministry of Labour and Social Policy (MLPS) which defines and ensures the essential levels of performance related to the vocational training system; the Autonomous Regions and Provinces with exclusive jurisdiction on vocational education and training both in terms of programming and management and delivery of educational offer; the Social Partners, which help to define and implement active labour market policies, particularly in the field of vocational training" (Franzosi, 2016, our translation).

The first repertories that were able to identify and release vocational qualifications that contained descriptive skills were the regional ones and date back to 2004, year in which a Regional Qualification System was developed [13]. The regions concerned themselves with vocational training aimed at achieving certain skills and worked together to define these systems. The regions were indeed responsible for vocational training aimed at achieving certain skills and worked together to define these systems. To operate in this area, it was necessary to bring together the needs and interests of different actors and stakeholders. The institutions and the actors that have collaborated for the identification and the sharing of the solutions have been (Armaroli, 2007): (i) the Region institution, typically expressed by the Regional Council, the Department of Reference and the competent Regional Service, which has the function of address, regulation and control of the Regional System of Qualifications; (ii) the Labour Market, expressed by employers, employees and their representatives, public and private entities providing services for employment, that expresses, requires, uses and intermediate professional skills; (iii) the Vocational Education and Training System, expressed by public and private entities providing vocational education and training services, which forms and develops the professional skills required by the labour market.

The Regional System of Qualifications has been identified and defined with the Social Partners and the professional profiles of which it is composed are to "broadband", that is, broad figures, who prefigure (potential) competences that express themselves and can express themselves in different agile roles. The professional profiles represent professional competences present in the economic and productive system of every region and are relevant for the specific regional policies of field of which the development is previewed [13]. Subsequently, in 2005, a Regional System of Formalization and Certification of Competencies was established, which provides for a rigorous evaluation procedure that allows to formalize and possibly certify the skills acquired by the person both in situations and educational and professional contexts, as well as in social and individual [13]. This was the origin of the Atlante del Lavoro experience.

Certification of competences and Atlante del Lavoro

The differentiation that we have seen created over time between regions poses various problems. There are quite a few cases in which it happened that two Italians from two different regions had obtained the same certificate for identical or similar qualifications with training courses of even considerably different duration. A minimum standard had to be defined (Ghergo, 2011). The solicitations deriving from the EU commission have placed the bases in order to resolve this problem. In order to comply with the guidelines of the European Commission and to have a single standard that would give a national reference and recognition of regional qualifications, allowing

[6] As is well known, the two ministries of education and university are now separate.

the mutual recognition of the qualifications issued by different regions, work began on a National Repertory of Educational and Training and Vocational Qualifications, also called Atlante del Lavoro e delle qualificazioni [13]. The Repertory provides a uniform framework for the certification of competences: it normat in the same framework the qualifications issued in the following areas of the national system of lifelong learning: school, university, vocational education and training, regional vocational training, qualifications acquired through an apprenticeship contract, the standardised and regulated professions [14].
In the section which constitutes the National Framework of Regional Qualifications, all Repertoires formally adopted by the Autonomous Regions and Provinces have been imported. In order to set it up, it was necessary to involve all the actors who have a role in the definition of professional qualifications. A technical group chaired by the Ministry of Labour and Social Policy and composed of representatives of the Ministry of Education, University and Research and the Autonomous Regions and Provinces of Trento and Bolzano was formed in 2013, with the technical support of Tecnostruttura and the Regions (Mazzarella, 2017).
The regional repertoires from which they started were very different from each other, this is because each region was specialized in some particular professional areas, and because the same professional profile could require different skills depending on where it works. The creation of a unitary reference for their correlation and equivalence and their progressive standardization was made possible thanks to the coordinated work of the different regions which divided the professional areas. Each of them worked on the professional areas in which it was most competent and for the operational description it unpacked the processes in areas of activity (ADA), in order to identify smaller elements that could be a unique reference (Mazzarella, 2017). From the association of regional qualifications to the ADA, and to the expected results, in fact, the table of equivalences or correlations arises which constitutes the basis for the national recognition of regional qualifications, for their certification and for the recognition of credits [14].
To solve the problem of the diversity of skills that can characterize the same professional figure, a compromise had to be found by addressing two contrasting tensions: On the one hand, the need to construct descriptives large enough to minimize their obsolescence, on the other, to take into account the characteristics that can distinguish the same profession according to the place and the period in which it is located.
This has created a structure able to give a unique and transparent national reference on which to build the training and learning processes and create a bridge with the labour market (Mazzarella, 2017). Created primarily to ensure national recognition of regional qualifications, Atlante performs other important functions in support of professional integration such as recognition of training credits, validation of competences acquired from experience, certification of competences acquired in diverse contexts. It is also an important tool used to support job orientation processes, guidance counselling, training planning and access paths to the labour market [14].

The challenge of building the repertoire has been addressed independently from other classification systems, for example no reference has been made to the European ESCO system, although it is articulated in the levels of the European Qualifications Framework for Lifelong Learning (EQF).
The situation described for the Italian regions is in fact analogous to that which occurs between the countries of the European Community and, more generally, between the various states. There are, for example, cases in which different denominations are used in the various states for similar professions or in which the same profession is viewed differently according to the productive, technological and professional contexts (Ghergo, 2011). Several attempts have been made to solve the problem and to try to harmonize training systems and taxonomies, through a series of directives, recommendations and highlighting of common denominators in order to achieve mutual

recognition of the qualifications issued by the various national authorities. Steps have been taken to meet this need, even if it has had to deal with a training and professional reality particularly attentive to national specificities and with marked cultural and structural differences qualifying the different systems that can hinder the homologation operations (Ghergo, 2011).
In recent years, also at European level, there has been a resumption of reflection on industrial policies and on the link between industrial and vocational training policies. In many European regions, public policies have tried to identify local production systems, sectors or supply chains on which to leverage to drive development.[7] The attempt by O*Net or Open Badges to reconcile classification needs with the ongoing change processes is perhaps the most advanced experience to date (cf. Appendix 2).

Vocational training and local knowledge (once again)

The fact that each Region interprets the certification of competences differently certainly has many concurrent explanations. It is, however, reductive to attribute these differences to a more or less competent administrative class or more or less attentive to European directives. At this level, local specificities also count, including skills and knowledge of place associated with the more or less widespread presence of certain productive activities and with the specialization of production systems located in each territory.
The theme of local knowledge, mentioned in the previous pages, comes back on the scene with all clarity. Even a quick and partial examination of some of the main characteristics of the Italian vocational training system clearly shows that attempts to arrive at an extended and shared system of professional declarations run into enormous difficulties in incorporating the specificities of places and their impact on knowledge. The point is clearly put forward by Saul Meghnagi in an essay of some time ago.
Every cognitive act, reminds us of Meghnagi, is a response to a set of specific circumstances. "Learning can […] be considered as a socially situated activity within the framework of a particular reality of work of residence, of life" (Meghnagi, 2005 p. 10, our translation). Professional competences are the result of a complex path that includes school education, and the experience gained on the job[8]. What matters is the specialization of the companies in which it has been worked, their dimension, the organizational models adopted, the autonomy of which it has been possible to enjoy in the performance of the own task and so on.

“Professional competence […] depends on many factors, related to what has been acquired in and outside the training system, in relation to the times and spaces in which one or more activities can be carried out. It is defined according to the margins and innovation allowed, to the forms with

[7] An example of this nature are the ClustER of Emilia-Romagna. The ClustERs are a public-private mixed-participation organization which has the task of guiding the process. In this role, public operators, universities and entrepreneurs also define the training paths and professional profiles considered key. The important point is that the figures identified are often completely new and can not be linked to the existing declaratory. In these circumstances, the problem facing competence classification systems is to find operating methods which, once again, in the presence of significant changes, will enable them to be incorporated.

[8] It is perhaps useful to remember that the theory of capital today looks above all at the scholastic and formative path. Many economists' studies, unfortunately, assume that a worker is qualified using only the number of years of school attendance as a proxy. Many refer to Gary Becker, but very few remember that the theory of human capital was founded by Jacob Mincer who looked primarily at work experience and on the job learning processes.

which it is possible to perform a given job, the models of authority and responsibility attributed or assumed" (Meghnagi, 2005, p. 10, our translation).

In the field of vocational training and certification of competences, as well, the understandable need for generality, which is ultimately instrumental in protecting the worker, conflicts with diversity and specificity of place that are difficult to overcome. Here too, Polanyi's paradox manifests itself with all its force.

8. Summary and conclusion

In the preceding pages we have told the complicated story of Prime and Second and the questioning and sometimes demolition work on the part of the Bespectacled and, to an even greater extent perhaps, the Rebel. When it comes to conclusions, it remains to pull the strings and understand what difficult task awaits Ultimo. Before trying to outline some of the possible scenarios in which Ultimo will have to operate, it is useful to make a quick summary of what has emerged.

Occupational and skills classification systems have a variety of functions in ensuring the proper functioning of labour markets. The occupational and skill classification systems are not just about the economists and the statisticians or the public decision-makers to provide essential information for policy design. They are necessary as well and especially to facilitate the matching between job seekers and employers, to support those who exercise a profession in one place and want to move to another and so on. In this way, an adequate classification system is, in the first place, a source of protection for the worker. In particular, in labour markets not subject to improper workers' intermediation the issue of competences' certification is crucial. The intervention of a third party is required with respect to the worker and the employer: a party that acts as guarantor, with great credibility and reputation.

There are therefore many reasons - related to comparability/portability, recognition of skills and, ultimately, job protection - why a variety of institutions in individual countries and on a supranational scale are striving to achieve homogeneous classification systems.

These needs, however, conflict with a wave of technological change of unprecedented scope and speed. The changes in technology are compounded, often in a wider variety, by the specific features of places, including the specific features that technical progress takes on in different production systems.

Both aspects contribute, in different ways, to undermining the foundations of classification systems that are updated very slowly, often as a result of agreements between different governments and different institutions. The characteristic aspect of these transformations, as has been argued, is not the induced disappearance of trades/professions. It is, on the one hand, the change of roles and functions associated with technical progress and, on the other hand, the translation of these roles and skills into specific industrial cultures and local training systems.

The increasing attention that in recent years has been paid to occupations and skills (and therefore also to the methods of identification and measurement related to them) was prompted precisely by and radical nature of the processes of change in progress. And yet, precisely for this reason, they should be used with full awareness that classifications provide us a partial and, to some extent, blurred image. Classification systems are in fact influenced by two opposing forces: on the one hand by digital transformation or other technological, economic or social tsunamis that lead to managing

change through a push for generalization, on the other hand from the solicitations that arise from specific production systems at local level. In addition, the fast and changing environment that characterizes the labour market today makes it even more difficult to define a coding standard in step with change. This is particularly true in relation to skills. “The competence […] can be indicated conventionally by reference to a professional figure or profile, described through a contractual declaration […]. However, it may be difficult to be clear if reference is not made to the particular characteristics of a context of action and to its possible explication in other areas” (Meghnagi, p. 224, our transaltion). The context of time and place matters. The lesson of Michael Polanyi, adopted by David Autor to interpret the effects of digitalisation by Italian scholars of local development (first and foremost Giacomo Becattini and Sebastiano Brusco) allow us to reconstruct the puzzle proposed in these pages.

In Italy, the change in the second part of employment contracts, exemplified by the metalworkers' contract, the persistence of strong differences in regional educational systems, the differences between regional systems of validation/recognition of acquired competences (both informally and through specific training paths) are evidence of problems that the available taxonomies fail to address adequately. At an international level, the Open Badges-based certification system in which ESCO (and its member countries) is working is emblematic of the high demand for flexibility in defining/recognising emerging occupations and professional profiles. It is, conversely, further confirmation of the insuperable limits of static systems of identification/description of skills and professions in times of rapid change.

Complicating the picture is the fact that, as has been argued, many of the classifications in use are also the result of a Fordist and Taylorist type of work organisation (and industrial relations system).

A static system of classifications and competencies, in fact, lends itself easily to describe a Fordist work organization, in which tasks are rigid, as proceduralised as possible and as invariant as possible over time. Much less so when the contours associated with emerging professional profiles are more blurred, varied, and associated with variable mixes of "hard" and "soft" skills.

What has been argued in these pages has profound implications for both training systems and industrial relations.

If the direction of change induced by digitization is as indicated in the preceding paragraphs, there is an inescapable problem of continuing vocational training for a very large portion of workers. The training will have to be linked to the way in which the new technologies take shape in the different business realities. It cannot be done in pills, in the abstract, with catalog courses. And it is equally evident that it must be much more capable than it is today of enhancing the specificities and knowledge of the place.

The implications for the industrial relations system are even more relevant. In labor markets in which the protection of the worker is reasonably well established, the developments mentioned above will necessarily lead to very flexible decentralized regulations, left to the parties (and on which the labor judge has great difficulty in intervening). This does not mean deregulating, but regulating with recognition of experience and certification.
But in markets where the system of protection is weak (as happens in many markets where migrant labour prevails) the lack of mechanisms for the recognition and portability of acquired skills (as in fact already happens in many personal care services - and not only) will only multiply the extension of Brainwaste phenomena.

New avenues may open up on the ground of job descriptor design. Perhaps the technological evolution itself could provide a way out. In order to identify new occupations, the majority of study rely on surveys, employer interviews, trade publications, job postings and the corresponding job titles, in addition to the current occupational classification. Similarly, new skills can be identified through tools such as surveys, interviews, skill classification and case studies, or by forecasting and anticipating the skill needs of the future (Beblavy *et al.*, 2016). However, as noted by Kilhoffer (2020, p.4), "These methods tend to rely on outdated or irregularly updated data, data focused on a specific case or sector, or derived from the opinion of an expert or stakeholder, and therefore difficult to generalise. To address these issues, identifying new occupations using more recent and representative data would be useful". The author proposes to use online real-time labour market data through the application of web scraping techniques.

These findings suggest that an automatic data-based methodology is needed to create classifications systems that are based on real-time informations that constantly update the databases. Companies such as Google or LinkedIn have developed a machine learning classifying method to continuously update their ontology in order to enter the job search services market (Hernandez, 2018). They both started from O*NET ontology, and they have developed then machine learning techniques to classify user entries on their platform. LinkedIn algorithm learns from users' response to suggestions made by the system and is in this way constantly updated. These examples suggest how the use of computer science and data analytics can be used to complement the traditional ontologies to provide real-time informations which are closer to showing the real evolution of the labour market (Hernandez, 2018).

As storytellers, the authors do not have a conclusive solution. Ultimo will find himself part of a world where Prime, Second, Rebel and the Bespectacled will continue to exist, often they will go their own way, sometimes they will meet and influence, help or clash each other. Perhaps they will reach a balance when Prime and Second, who capture the static, schematic, and partial aspect of reality, will collaborate with Rebel and the Bespectacled, which represent the characteristic elements of the context of action. None of them can in fact, alone, come to represent reality. Each of them focuses on its own particular aspect and, only by collaborating together, will paint the world with as many colors as possible.

Appendix 1. Occupational classification systems: US, EU and Italy

Definitions, methodology and structure

Occupational classification systems are developed in a variety of ways. One commonly used qualitative development method is known as the top-down approach. Such systems often have two or more hierarchical levels in which to group and are developed by experts who have knowledge of the occupation or job information. The information is grouped into a structure of more generalized occupational groups, using characteristics such as work function, job title, or skill level. Because the developers usually tailor classification systems for a specific purpose, there is no consistency in the number of hierarchical levels or number of job titles included.
A second empirical approach involves applying statistical techniques such as cluster analysis or factor analysis to the knowledge, skill, and ability data associated with various jobs. The job analysts name the resulting clusters that become the hierarchical levels within the structure. These

classification groupings are based on data rather than subjective opinion, which can be advantageous for purposes such as validation and research. However, classification systems developed using only statistical methods can be difficult to interpret and may lack the face validity necessary to be widely accepted.
To address this issue, a third approach combines the qualitative and quantitative approaches described above. For example, experts design and develop a structure of occupations using the top-down approach, based on some need or organizational objective. Once the hierarchical structure has been established, statistical analyses are used to verify, modify, or provide validation support for the rationally derived hierarchical classification schema [1].

There are defining elements proper to the different classification systems that must be made explicit. In particular, the definitions of occupations / professions and skills and qualifications.

Occupations and jobs are distinct elements in an occupational taxonomy. A job is defined as “a set of tasks and responsibilities performed by a person, for an employer or for oneself (International Labor Office” (ILO) 2012). While an occupation is “a set of jobs that are carried out, with slight differences, in multiple establishments, and not necessarily within the same industry” (Emmel and Cosca 2010, Ospino Hernandez. 2018).
The terms ‘skill’ and ‘competency’ are used interchangeably in many contexts, and they are similar but differ in important ways. The European Qualification Framework (EQF) (European Union 2008) define a competence as “the proven ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development” and it is described “in terms of responsibility and autonomy”. A competency is therefore any expertise or talent that is useful for a job. Examples of competencies can vary widely but include developed capacities (e.g. active listening), proficiency with tools or technology (e.g. lancets, Microsoft Word), innate abilities (e.g. originality), and academic knowledge (e.g. medicine).
Knowledge is defined by the EQF as “the outcome of the assimilation of information through learning. Knowledge is the body of facts, principles, theories and practices that is related to a field of work or study.” Finally, EQF defines ability (skill) as “the ability to apply knowledge and use know-how to complete tasks and solve problems.” Referring to something as a skill, depending on the audience, may denote different subsets of competencies, or just as a shorthand for any competency (Crocket *et al.*, 2018).

The definition of qualification is complementary. Qualification is defined as “the outcome of specific training, education, work experience and shows a significant interdependency with the personal attributes of an individual. Qualification summarises knowledge, skills, and capabilities which are required by specific activities of a job or daily life. From an individual’s point of view, qualification is a precondition for successful occupation and job fulfilment, because the status of development influences his/her market opportunities and thus his/her labour market value” (Cedefop 2006).

***

In the following pages the Italian national system of occupational classification, “Atlante del Lavoro”, is compared with the american system O*NET (Occupational Network Information) and the European system ESCO (European Skills, Competences, Qualifications and Occupations). First, the macro characteristics of the systems are presented and then they are compared.

O*NET

O*NET (https://www.onetonline.org/) was launched in 1998 by the US Department of Labour as an online version of the Dictionary of Occupational Titles (DOT), developed more than 50 years ago. O*NET data inform of important activities in workforce development, economic development, career development, academic and policy research, and human resource management. The data are organised as a content model with six domains. One important advantage is that it is updated very regularly: a new version of the O*NET database is usually published annually in late June. The characteristics of about 750 individual occupations have remained quite stable, and they every year 100-120 occupations have been updated.

The two O*NET core elements are a content model and an electronic database fed by a data collecting programme.
The content model provides a framework for more than 400 variables describing about 1.100 occupations based on the SOC.
The descriptors are organised into six major domains, which enable the user to focus on areas of information that specify the key attributes and characteristics of workers (the first three domains) and of jobs (the last three domains), and are either crossoccupational or occupation-specific (Cedefop 2013).

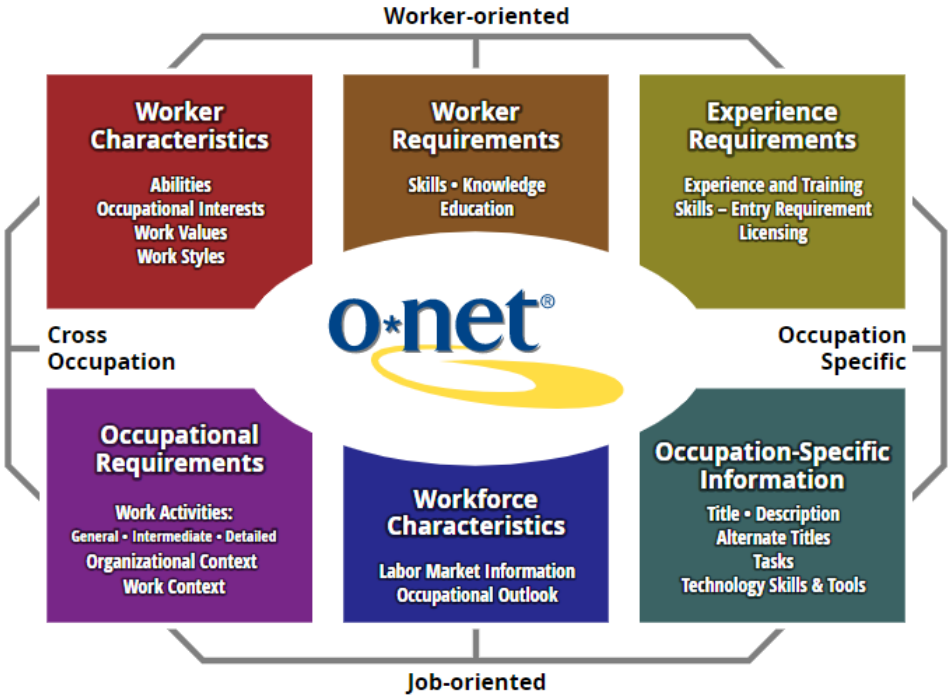


*Fig. 2: The O*NET content model. Source: O*NET resource center, http://www.onetcenter.org/content.html.*

ESCO

ESCO (https://ec.europa.eu/esco/portal) is a valuable tool linking skills and competences to occupations and bridging the information gap between the worlds of work and learning. It is organized in three pillars interrelated with each other: the occupations pillar; the knowledge, skills and competences pillar; the qualifications pillar. ESCO further aims to contribute to labour mobility, online matching and shifting labour outcomes. ESCO's occupation pillar is linked to ISCO (Beblavy,

et al. 2016). It has been developed in all the member state's languages of the European Union (Crockett *et al.*, 2018).

In reconstructing the framework of the US and European models, the guidelines of the OECD deserve particular attention. The broad OECD Skills Strategy is outlined in an important document entitled "Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies" (OECD, 2012). The strategy comprises three pillars (Fig. 3). The first, "Developing relevant skills" aims to arrive at a skills supply of a sufficient quantity and quality. The second, "Activating skills supply" aims to re-integrate inactive individuals into the labour force to ensure that all available skills are used. The third, "Putting skills to effective use", is focused on skill-matching.
Importantly, the OECD reports that new skills often are developed informally (e.g. through work experience). Moreover, the OECD is also concerned about the deterioration of skills that are not put to use[9]. Throughout the report, there therefore is a strong emphasis on life-long learning [24].

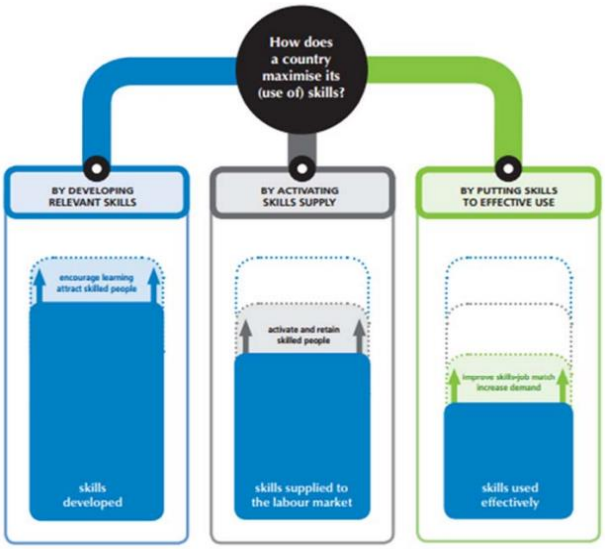


Fig. 3: The OECD skills strategy framework

Atlante del Lavoro

The "Atlante del Lavoro e delle Qualificazioni" has been online since the end of 2016 (http://atlantelavoro.inapp.org/). It is the main informative tool of titles and qualifications on the

[9] PIAAC is an initiative of the OECD (Organisation for Economic Cooperation and Development). The "Survey of Adult Skills" is part of the Programme that assesses the proficiency in literacy, numeracy and problem-solving skills (in technology-rich environments) of adults in the context of their socio-economic status (for 33 countries). It evaluates the availability of these skills and their use at work and at home. In this way, it provides valuable information for educators, policy-makers and labour economists. With the survey, the OECD aims to support the development and implementation of national skills strategy. The results of the survey are presented in the OECD Skills Outlook. The survey reveals that more education is not necessarily associated with better skills, and that skill acquisition beyond formal education (at work or at home) is becoming increasingly important. The 2015 report further recommends to ensure that all young people level school with a range of relevant skills, to assist school leavers to enter the labour market, to dismantle the institutional barriers to youth employment, to identify and help young people who are not employed or in education to reengage, and to facilitate better matches between young people's skills and jobs.

national repertoire. It consists of a detailed map that describes the world of work and qualifications. It describes the contents of the work in terms of activities (job, tasks,…) and potentially deliverable products/services in performing the same described activities. The contents of the work are represented, and made navigable, through a classification scheme consisting of 24 professional economic sectors (SEP).

The contents of the work are described in a process perspective and the activities are defined up to their minimum level, allowing the allocation of the individual qualifications, contained in the regional repertoires, in the ADA (areas of activity), making them comparable.
Atlante's main function is to ensure the national recognition of regional qualifications, but it can perform other important functions such as the recognition of course credits, the validation and certification of competences. It can also be a valid support for training planning, for the planning of access routes to the labor market, for continuous training for development and retraining or professional retraining [2].

Comparison and insights

With the purpose to make a comparison between the systems presented, first the comparison metrics are defined. Then, more detailed information on the systems have been sought, mainly consulting the technical reports present on the official websites of the systems [3][4][5]. In addiction, because Atlante del Lavoro is still under development, an interview with one of the creators has been conducted in order to find more detailed informations about the system.[10]
In the following pages, the insights that emerges from the comparison are presented.

Structure

O*NET is aligned with US standards, ESCO with the European. ESCO is structured according to an ontology, while O*NET does not, it has its own model of definition of the categories and the relationships between them. An ontology is defined as a knowledge graph that allows to limit complexity and organize information into data and knowledge, which includes a representation, a formal naming and a definition of categories, properties, and relations for certain knowledge domains (Crockett *et al.*, 2018). This system helps to identify the skills and qualifications related to an occupation and allows management of the workforce relying not only on job titles and duties, but also on the required skills. This is particularly important because in a changing labor market, tasks performed by workers can shift with technology adoption (Ospino Hernandez 2018).

ESCO describes indeed the three pillars in which data on occupations, skills and competence and qualifications are organized: the occupations pillar; the knowledge, skills and competences pillar and the qualifications pillar. Occupations are explicitly related to skills, for transversal skills there is information about how they are related to each other.
Atlante del lavoro also follows an ontology. It is hierarchically organized, it starts from the set construction, the second layer is the economic-professional sectors, the third is the process sequences, the fourth is the Areas of Activities (ADA) and finally the last contains the activities that aggregate the expected results. The ADA corresponds to "a meaningful set of specific activities,

[10] An excel table has been made to summarize the information found. It contains the systems on the lines and the comparison measures on the columns. The table can be consulted by clicking on the following link: https://docs.google.com/spreadsheets/d/1Si3r_Y-vr7tUrVT9uz2fZ6xi__u4_HlqTTVn06Gq7Xo/edit#gid=0.

homogeneous and integrated, result-oriented, and identifiable within a specific process. The activities which together constitute an ADA are homogeneous both for the procedures to be applied and for the results to be achieved and, finally, for the level of complexity of the competences to be expressed" (Perulli, 2013).[11]

Classification system

O*NET expands the taxonomy of the Standard Occupational Classification (SOC-2010) developed by the U.S. Bureau of Labor Statistics. For occupations in it, O*NET uses the basic 6-digit numerical coding structure of the SOC as its framework and adds two digits at the end of each SOC occupation to differentiate unique O*NET occupations within the SOC system with the aim to preserve or increase the level of detail in an occupational category. In order to differentiate unique O*NET occupations within the SOC system if the level is kept, the digits 00 are added on, whereas, if a SOC occupation is expanded, then the digits start at 01. The latest version of the O*NET taxonomy matches the SOC 2010 structure (Ospino Hernandez 2018).
ESCO is related to the EU national classification ISCO-08, developed by the ILO. ESCO occupation is mapped to one ISCO-08 unit group, the two classifications are interoperable. This allows ESCO to build on the international acceptance of ISCO. This is particularly important because most national occupational classifications are currently mapped to ISCO-08. This will also make it easier to map them to ESCO. Some EU Member States developed and currently use occupational classifications to deliver labour market services at national level. The Commission services used several of these classifications as reference during the development of ESCO. Additionally, since ISCO-08 is currently used to enhance the international comparability of statistical data, it makes ESCO an interesting tool to support labour market statistical reporting (European Union 2019).
Moreover, ESCO is in relation with the EQF, a common reference framework originally adopted in 2008 that helps learners, graduates, education and training providers and employers to understand and compare qualifications awarded in different countries and acquired in different qualification systems in Europe. National databases of qualifications provide their data to the ESCO qualifications pillar and information on the EQF level of the qualifications they contain, therefore fostering transparency and comparability (European Union 2019).
There are some connecting tables that allows to switch from an occupation in ESCO to the equivalent in O*NET available at the following link: http://ibs.org.pl/en/resources/occupation-classifications-crosswalks-from-onet-soc-to-isco/.
Atlante del Lavoro is connected with the two main Italian classification systems which are the ISTAT classification of professions CP 2011, and ATECO 2007 and these two classifications in turn derive from international classifications, the CP ISTAT comes from ISCO which is also the classification base on which ESCO was built at the fifth digit. Connection with ESCO is possible through the CP ISTAT 2011. ATECO in turn derives from an international classification NACE which is a classification of economic sectors that has got an international value.

[11] All systems use a "key" to uniquely identify a database element. For ESCO any concept is identified by Uniform resource identifier (URI), for O*NET by an "element ID", for Atlante del Lavoro by a variable string and for PIAAC by a variable string of type "name".

Skills and competences

O*NET includes generic skills, interests, styles and working values, tools and technologies which are related to Occupations. O*NET, when referring to "skill", means the developed capacities (for example active listening).
ESCO instead contains a real skills mapping, but does not cover the skills, interests and styles and working values typical of O*NET. The terms "skill" and "competence" in ESCO are interchangeable and therefore some people might think that skills are a subset of competences (Crockett *et al.*, 2018). The knowledge, skills and competences pillar, also referred to as the "skills pillar", provides a comprehensive list of skills that are relevant for the European labour market.
Atlante del Lavoro does not contain the skills itself, but skills, ability and competencies are those of the education and training systems of Italian country. They refer to it as a website that also contains and is the framework within which all the qualifications and competences of Italy fit, because they are those of the qualifications that are connected.
An example is made in order to better understand the concept: Atlante del Lavoro has a process that describes the engineering design of a house, inside it are specified the work processes needed to make a home, connected to these processes there is a degree in building engineering necessary to execute them that has incorporated the skills that must hold a worker that has a degree in building engineering. Skills are therefore connected to but not part of Atlante, they belong to the qualification in Construction Engineering of the Italian University.

Occupations

In O*NET the Occupations are uniquely assigned to an ID and are often structured hierarchically, mapped according to the structure given by the "Standard occupational classification" (SOC), the standard statistical system used by the Federal Statistical Agency.
In ESCO they are structured according to the ISCO-08 mapping which provides the first four hierarchical levels on which ESCO is based to establish the next levels. The structure of ESCO is similar to that of O*NET, although it is much denser and deeper.
The occupations pillar aims to describe all occupations relevant for the European labour market.
In Atlante del Lavoro occupations are organized according to CP2011: the Official Classification of Professions realized by ISTAT. The CP2011 classification provides a tool to bring together all existing jobs in the labour market within a limited number of professional groupings, to be used to communicate, disseminating and exchanging internationally comparable statistical and administrative data on professions, which should not be seen as an instrument for regulating occupations (Mazzarella and Porcelli 2017). Since the CP2011 comes from ISCO, the occupational classification follows the structural updates of the international ISCO classification.

Qualifications

The qualifications are included in the O*NET Content Model that describes the characteristics of an occupation. Its hierarchical model consists of six domains, describing the day-to-day aspects of the job and the qualifications and interests of the typical worker. It does not have, however, a specific section dedicated to qualifications.
Qualifications in ESCO come from national qualifications databases of Member States and are included in National Qualifications Frameworks that have been referenced to the EQF. The Commission has been financially supporting Member States and other partner countries to develop national qualifications databases and to interconnect these with ESCO. Other qualifications that are

not part of national qualification frameworks but are also relevant for the European labour market might be provided to ESCO by awarding bodies in the future. The qualifications pillar aims to collect existing information on qualifications and, in contrast with the other two pillars, is populated only by external sources.
The third section of the Atlante del Lavoro, "Atlante e qualificazioni" brings together the qualifications awarded in the different areas of the lifelong learning system: school, education and vocational training, higher education and regional vocational training.

Relationship between the pillars

In O*NET technology skills are related to occupations.
In ESCO occupations are explicitly linked to skills, for transversal skills contain information on how the skills are connected to each other.
Atlante del Lavoro includes relationships between areas of activity (ADA) and qualifications. The methodology according to which activities are related to qualifications is a correlation link methodology. According to that, the entity holding that qualification and therefore the University, the region which holds the qualifications for vocational training and the Ministry of Education link, relate their skills and qualifications to activities. The activities act as the element with respect to which all the qualifications are compared.[12]

Multilinguality

ESCO is developed in all the member state's languages of the European Union, 26. Data are therefore transparent and readily available to different stakeholders (Ospino Hernandez 2018). This is important to support the labor mobility in Europe because every person in the European Union can upload the Europass curriculum in their own language and the system allows the employer to read the curriculum automatically translated by the system itself into the language of its own country.

Moving around or travelling from one country to another for work or study purposes has indeed become for many people a necessity or an aspiration for greater well-being and a better life. Half of the European citizens believe that the right to work abroad is a positive achievement, although according to the Eurobarometer survey only 2% of EU citizens of working age live in a European country other than their own (in the USA this percentage is 6%) (Perulli 2013). The O*NET system is originally built up in English and it has been developed a Spanish version. Atlante del Lavoro is developed only in Italian.

[12] The O*NET database is provided in five formats: Microsoft Excel (XLSX); Tab-delimited text file; SQL files for MySQL; PostgreSQL, or compatible relational databases; SQL files for Microsoft SQL Server; SQL files for Oracle Database.
The ESCO classification is currently available for download in three data formats: SKOS/RDF format: Full dataset with all concepts and relationships in all languages; works fine with Virtuoso triplestore; CSV format: Partial dataset with relationships or with concepts from one ESCO pillar in one language, e.g. for import into Microsoft Excel; ESCO API: Web based service that provides applications with access to the different versions of the ESCO classification. The functionality covers the majority of the ESCO business cases.
At the moment Atlante del Lavoro is available only navigable, therefore it is not possible to make extractions in the public area. Excel extraction is viable in reserved areas (e.g. entities holding the regions). In the future, more on less from the autumn, excel extraction will be made available also for the public area.

Problems related to these systems and ideas for improvements

The classification of occupations poses a number of specific problems.
The creation of these systems is time-consuming, costly and are not able to be adapted to the today's fast-changing labour market environment. The labour force is indeed constantly transforming: it is expected to become older and ethnically and racially diverse, due to the globalization's talent-migration tendencies (Morkowitz). Moreover, technological innovation will eliminate some occupations and create new ones. Finally, new laws and regulations will be implemented, and customer tastes are continually changing (Morkowitz). In general, therefore, it seems that in the future the labour market will be different from today's. An example of these evidences is shown by a Business insider article [9] that states that "65% of children that are starting the elementary school will do a job that doesn't exist today".
Even if each system is periodically reviewed and renovated, they lack a continuous update process in real time. Such static (or little dynamic nature leads these systems to be out of date and always late in representing the evolution of the labour market and the future jobs and to represent occupations that are no longer exercised (Leeuwen, 2004).
Another problem is that new and old professions are described in the same way, but more explanation is needed for those new job's profile that are less known.
The systems omitted seasonal employment and illicit or semi-legal economic activities [12].
These findings suggest that an automatic data-based methodology is needed to create classification systems that are based on real-time information that constantly update the databases. Companies such as Google or LinkedIn have developed a machine learning classifying method to continuously update their ontology in order to enter the job search services market (Hernandez, 2018). They both started from O*NET ontology, and they have developed then machine learning techniques to classify user entries on their platform. The LinkedIn algorithm learns from users' response to suggestions made by the system and is in this way constantly updated. These examples suggest how the use of computer science and data analytics can be used to complement the traditional ontologies to provide real-time information which are closer to showing the real evolution of the labour market (Hernandez, 2018).

Appendix 2. - Skills certification systems: US, EU and Italy

In 2016 the European Commission adopted the New Skills Agenda for Europe, one of the primary initiatives in the Commission Work Programme (Benadusi and Molina 2018). One of the key priorities is making skills and qualifications transparent and comparable throughout Europe.

Qualifications are useful for employers for understanding what people know and are able to do. However, relying only on these makes it difficult to capture skills acquired outside formal learning institutions, such as work-based learning and experiences abroad [3], wich therefore risk being undervaluated. Identifying measuring and evaluation methods and tools for the certification of skills gained in non-formal and informal contexts is becoming particurarly important, expecially for people with lower qualifications, the unemployed or those at risk of unemployment, those who need to change career path and for migrants. It helps to focus the investment on the necessary vocational education and training policies and skills development, to identify further training needs

and take up opportunities for re-qualification, to improve people's employability using their experience and talent, and help to reduce the current gap and enable a smooth transition in the world of work [3]. According to the periodic monitorings carried out by Cedefop, about half of the EU Member States have already developed and adopted a formalised strategy for the validation of expertise obtained from experience, while the others, including Italy, are still developing or testing such a strategy (Perulli, 2013).
The occupational classification systems that we have presented and compared in the previous section play a key role during skills assessment and certification. They formalize a common language based on wich occupations and qualifications are described. This allows to emphasize the role of the validation of competences for the acquisition of a recognized degree or qualification and therefore to encourage the transfer and accumulation of skills, increasing flexibility and enabling individuals to undertake new paths of learning.
In the section that follows, certification process will be discussed in more detail. First, a definition is given. The different kinds of learning outcomes and a general overview of the process will be presented, then the paper focuses on the specific countries, highliting the different processes and actors involved, and how the occupational classification systems presented in the previous sections are involved the process of identification, validation and certification of the competences.

Cedefop (2009) provides the essential definitory framework.

- *Certification of learning outcomes*: "The process of formally validating knowledge, know-how and/or competences acquired by an individual, following a standard assessment procedure. Certificates or diplomas are issued by accredited awarding bodies."
- *Validation of learning outcomes*: "The confirmation by a competent body that learning outcomes (knowledge, skills and/or competences) acquired by an individual in a formal, non-formal or informal setting have been assessed against predefined criteria and are compliant with the requirements of a validation standard. Validation typically leads to certification" and also as a "Process of confirmation by an authorised body that an individual has acquired learning outcomes measured against a relevant standard."

"Validation consists of four distinct phases: 1. identification through dialogue of particular experiences of an individual; 2. documentation to make visible the individual's experiences; 3. formal assessment of these experiences; 4. certification of the results of the assessment which may lead to a partial or full qualification." (Cedefop 2009; Council of the European Union, 2012).

What is validated and certificated: formal non-formal and informal learning outcomes

Formal learning is provided in an organised and structured context (for example, in an educational or training institution or at work), specially designed for this purpose (in terms of learning objectives and time or resources provided for learning). Formal learning is intentional from the learner's point of view. It usually gives rise to a certification [41].

Non-formal learning is semi-structured and delivered within planned activities not specifically designed as learning (in terms of objectives, timing or learning support). Non-formal learning is intentional from the learner's point of view. The results of non-formal learning can be validated and can lead to certification.

Informal learning results from daily work, family or leisure activities. It is not structured in terms of learning objectives, time or learning resources. In most cases, informal learning is unintentional

from the learner's point of view. The results of informal learning can be certified if validated (Isfol 2015).

The validation of non-formal and informal learning guarantees specific benefits such as: economic benefits (reduction of the costs attributable to formal training as validation reduces the costs of acquiring a qualification, reducing the time taken to access to the labour market by high qualified people), educational benefits (self-esteem development and awareness of own's skills and lifelong learning approach), social benefits (validation increases opportunities for fair and transparent access to civil rights, education and the labour market for the most disadvantaged, unemployed young people and older workers who have not had adequate education and training opportunities), psychological benefits (increased self-esteem and individual motivation) (Perulli, 2013).

In companies' employee appraisal elements that, a few decades ago, were completely negligible become important. Companies begin to distinguish with awareness between different "types of skills". Among the different dimensions of the worker's competences/skills, those that are conventionally identified are the following (Cedefop, 2014; Orpinas, 2010, [8]).

-Profession-related skills and competences: also called "hard skills". They are related to working processes and resources. These are the most frequently and most extensively assessed.
-Social competences: is defined as the ability to handle social interactions effectively. Examples are leadership, managerial skills, and customer orientation or communication ability.
-Personal competences: are personal traits and abilities that affect your results in the workplace and in life. Personal competencies include self-awareness, drive, relationship skills and confidence. Personal and social competences are also called "soft skills" and are more difficult to assess than the previous. However, as we have seen in the earliest paragraph, they have lower risk of obsolescence.
-Digital literacy: with the digitalitation of work processes, this competence is becoming increasingly important, especially in administrative (office workers, clerks, etc.) and IT jobs.
-Language skills: are related both to foreign and mother tongue languages and have a great importance in management position and in those jobs who require a customer contact (sales personnel). Additionally, these are suitable to be tested in people who work in international environments and in migrants.
-Analytical-mathematical competences: are particulary decisive for accountants, bookkeepers and similar jobs as well as managers and engineers.

The american context and O*NET

The model in which O*NET was developed makes use of descriptors that highlight the characteristics of occupations (job-oriented) and people (worker-oriented) and is therefore suitable to be applied to all jobs, sectors and industries (Simoncini, 2016).

To understand how the American labour market is regulated, we need to introduce the so-called "Occupational Licensing" mechanism: a form of labour market regulation by which the US Government establishes the skills required to carry out an activity, in order to ensure consumer safety and an adequate level of service quality.
It provides for three different forms of attestation: the first is the registration, by which people register with government agencies indicating their personal data and qualifications. The second is the certification, that allows any person to perform his duties after passing an examination, at the

conclusion of which a certificate attesting the level of skill and knowledge demonstrated is issued. Finally, the licensure (the professional license) gives the right to practise a profession (Simoncini, 2016).
While producing a slight improvement in services, occupational licensing imposes net costs for both society and those who intend to certify their profession. Moreover, in the USA there is no single legislation for the whole territory, and therefore the risk of damage to the worker's mobility is high. O*NET's analysts then have the task of defining constructs within the information categories to develop a classification methodology that can be extended to all existing professions in order to create a common descriptive language.
The two american systems, O*NET and Occupational Licensing, are methodologically valid models as they both aim to protect the worker's professionalism, based on his skills: with the first one it is possible to understand which are the key requirements of a professional activity, with the second the standards to which to comply in order to exercise it.
There is a similarity between the American case and what is happening in our country. It highlights the tension between the efficiency of the system in terms of the technological infrastructure and the methodological system on the one hand and the diversity of the realities within the country that makes it difficult to accommodate every change on the other hand (Simoncini, 2016).

The european context and ESCO

The 2012 European Parlament Council recommendation on validation emphasizes the importance of constituting ways of validating non-formal and informal learning and obtaining obtain a full qualification or, where appropriate, a partial qualification, based on the validation of non-formal and informal learning experiences. The Member States are invited to linking the validation modalities to national qualifications frameworks in line with the European Qualifications Framework and align qualifications or the parts of qualifications obtained through the validation of non-formal and informal learning experiences with the arranged standards, which are equal to or equivalent to the standards of qualifications obtained through formal education programmes. The member states are at the same time called to take into account national, regional and/or local needs and specificities as well as sectoral ones (Isfol 2015).

The level of development and implementation of non-formal and informal validation systems in the Member States appears, at present, to be rather differentiated and heterogeneous. In some countries such as Denmark, France, Norway, Finland, the United Kingdom, Spain, Portugal and Iceland, the level of formalisation and implementation of the validation process is particularly developed and consolidated. It includes specific rules, laws and formal acts which introduces the procedure within the education and training and labour market systems. In other countries such as the Czech Republic, Poland, Germany and Hungary, the validation of non-formal and informal learning is still at an experimental stage and it is often entrusted to the initiative of employers' associations, trade unions, enterprises or third sector bodies rather than training or educational agencies through European programmes aimed at developing experimental models and good practices such as the Leonardo programme (Italy, Estonia, Slovenia). Denmark, Switzerland, Sweden, Poland, Norway and the Netherlands are applying validation mechanisms closely related with the labour market to cope with the economic crisis and rising unemployment. These countries have got rather consolidated technique of validation of non-formal and informal learning that are integrated with the socioeconomic system of reference. Many other countries are concentrated on the systems for validating adults learning (Slovenia, Romania, Portugal, Lithuania, Ireland, Belgium, Bulgaria). France, United Kingdom and Finland have anchored the system of validation of non-

formal and informal learning in structured devices for the recognition of skills in relation to the acquisition of a professional qualification. Furthermore, the federalist government systems structure in some of the Member States (Germany, Spain, Belgium) has made the validation approach very heterogeneous and differentiated, with diversified effects on the implementation of the devices and therefore also on the level of sharing by the public opinion and on the production systems of reference.
Overall, there seems to be evidence to indicate that the concept of validation of non-formal and informal learning has generally spread across all European countries, although the levels of formalisation appear to be heterogeneous (Perulli 2013).

As previously explained, ESCO is developed in all member state's languages of the European Union. This is extremely important because facilitates cooperation between countries and supports the mobility of learners between countries and systems. It also helps to understand the developments in skills and competences in an international context and this enables the vocational education and training to respond directly to the needs of the labour market. The standardized terminology in wich ESCO is developed allows to describe how occupations, knowledge, skills and competence (in which learning outcomes are commonly defined), and qualifications are linked and iteract with each other. It facilitates the dialogue between labour market and education and training stakeholder within and across sectors and borders. It also helps to describe, analyse and understand the relasionship between education and training and the labour market. ESCO terminology helps to identify, document and assess the outcomes of non-formal and informal learning and thus to support its validation and to underpin descriptions of assessment criteria and thus directly support qualification bodies at international, national and sectoral level (Europen Commission 2013).

ESCO is trying to define a method of certification that consists of Open Badges, even if is so far a side-project [6]. Open Badges are "visual tokens of achievement, affiliation, authorization, or other trust relationship sharable across the web. Open Badges represent a more detailed picture than a CV or résumé as they can be presented in ever-changing combinations, creating a constantly evolving picture of a person's lifelong learning" [7]. An ESCO digital badge is a visual representation of a soft or hard skill. It's available online and contains metadata which describes the badge: its meaning, origin and destination. An example of user that can benefits from it could be an employer, who wants to certify that his employee has a certain skill. He could use a badge to attest it [6]. The scope of ESCO badges was to give to non-IT stakeholders a user-friendly tool to create and issue visually attractive, ready to use badges while linking them with the ESCO taxonomy, as a way for the badge issuers, users and consumers to commonly understand the skills and competences the badges represent.
The idea is that when workers move from one organisation to another, the organization in which they have worked could give them a badge which states what occupation did they represent (via ESCO Occupation), duration of the experience and which skills they acquired during this time (via ESCO knowledge, skills and competences). The employee can then use this badge in their online portfolio and social network profile while seeking new opportunities.

The italian context and Atlante del Lavoro

In Italy there has been for many years a rich debate and a substantial agreement between all institutions and social actors on the importance of being able to validate the acquired learning in non-formal and informal contexts. However, the necessary provisions for the development and

institutionalisation of a national system for the validation and certification of skills acquired have not yet been adopted. Only at regional level have been formalized and implemented institutional systems and methods for the validation of non-formal and informal learning (Perulli, 2013). In recent years in fact, all the Italian regions, which are the main hub of work and vocational training services in the territory, are addressing the issue of certification and validation of skills within their work system or vocational training, contextualizing and differentiating tools and approaches.

Some regions are at an early stage of strategic approach to the theme, which is addressed by small steps, perhaps starting from specific supply chains and training types. In Abruzzo, Calabria, Campania, Friuli Venezia Giulia, Molise, autonomous province of Bolzano and Sicilia the strategy is not yet formalised and standardised and validation is applied within specific projects, programmes, types and training supply chains. Other Regions have instead arrived at a formalization of the strategies through specific deliberations and normative acts that include the validation within the regional certification system even though they have not yet implemented these programmatic indications. In Basilicata, Lazio, Liguria, Marche, autonomous province of Trento, Puglia and Sardegna a strategy has been formalised and standardised within a regional system of validation and certification but has not yet been implemented. Other regions, after having formalized and standardized specific devices for the validation of non-formal and informal learning, have also started concrete actions aimed at setting up the system also by "testing" gradually in the field the validity of the strategic and methodological approaches defined. In Emilia Romagna, Lombardia, Piemonte, Toscana, Umbria, autonomous region of Valle d'Aosta and Veneto the strategy has been formalised and standardised within a regional system of validation and certification and field activities related to these standards have been carried out [3].
However, it has not yet been adopted the necessary provisions for the development and institutionalisation of a national system of validation and certification of competences [4].

Process stages	Non-formal and informal learning		Formal learning
	Identification and validation	**Certification of competences after validation**	**Certification of competences acquired in a formal context**
Identification	Identification of skills, reconstruction of the experience and preparation of a dossier that documents the evidences.	Admission via "Validation document" or Validated Dossier	Admission through formalisation of the achievements of learning outcomes.
Evaluation	Technical examination of the dossier and possible direct evaluation (e.g. structured technical interview).	Summary evaluation carried out with structured technical interviews and/or performance tests. Presence of the Commission or of a collegiate entity that ensures the compliance with the principles of third-party, independence and objectivity of the process.	Summary evaluation carried out with structured technical interviews and/or performance tests. Presence of the Commission or of a collegiate entity that ensures the compliance with the principles of third-party, independence and objectivity of the process.
Attestation	Preparation (and possible release) of the "Validation document" or the Validated Dossier.	Certificate issue.	Certificate issue.

Fig. 4: Stages for validate formal, non-formal and informal learning in Italy. Elaboration from Isfol (2015)

The descriptive sequence in which Atlante del Lavoro is articulated presented in the previous chapter allows to identify the production processes of goods and services, and of the single activities that compose it. This allows to have a national unitary reference for regional qualifications, which makes it possible to verify and compare the skills and profiles described in the different regional repertories, which we have seen to be different in each region, and to have a parameter on which to evaluate performance. On 22 January 2015, an agreement established that regional qualifications which, in terms of competence, have the same work activities in a correlation group, are automatically considered equivalent to each other. Thus the national qualification framework is a single reference at national level for correlating similar qualifications and for identify, validate and certificate the competences (Tecnostruttura, 2015).

At a national level, all the regions and the public administrations ensure the constant participation in the work of the Technical Group Regions and the Ministry of Labour and Social Policy. They have the support of Isfol and Tecnostruttura, which is working at a technical level with the aim to define the methodological framework for implementing the indications expected by the Legislative decree 13 of 2013 (national system of certification of competences).

At a regional level, all regions are committed to have its own repertoire of professional profiles/qualifications and to adopt its own rules on the services of identification, validation and certification of competences.
Experts in the field are committed to make verifications on the reconstruction of work processes, ADA and activities carried out at national level. Futhermore, regions and public administrations have elected their own delegates that have the role to provide a database useful for the process of identifying correlation groups, which will be created to establish qualifications that are equivalent to each other and thus automatically recognisable at national level, on the basis of the work activities. All the regions and the public administations have thus developed a document about the system standards in the field of identification, validation and certification; have developed an association between skills of their qualifications and national ADA activities; have built up a methodological document for the correlation process, useful for the creation of correlation groups, and ultimately to establish qualifications which are equivalent in relation to the work activities (Tecnostruttura, 2015).

The process of validation follows the phases presented before. Figure 4 summarizes the Italian process.

Problems related to these systems and ideas for improvements

In the previous chapter, we have seen how the competence certification systems turn out to be useful in order to address the numerous issues connected with the ongoing socio-economic, organisational and technological changes. However, the certification systems pose a number of problems similar to those relating to the classification systems.

Like classification systems, certification systems are rigid too and are therefore not suitable to keep up with the transformations. A survey carried out by INAAP (Franceschetti *et al.*, 2019) on 550 thousand Italian companies illustrates this point clearly. The results have shown that about one-third of them demonstrate the need to update the skills of at least one of their employees.

Another problem is that employees may have acquired skills abroad that may not be present (or that could be mapped differently) in the certification systems of the country where they would like to work. The globalization's talent-migration tendencies require an international comparability of competences, in order to help workers to find more easily employment opportunities in new occupations, sectors or locations (Ernst *et al.*, 2018).
Lot of migrants have acquired skills through formal education and training systems in their own country which are not recognized in the state of destination could find themselves at disadvantage. This phenomenon, called *"Brain Waste"*, explains why a lot of migrants occupy unskilled jobs despite having higher qualifications. A study on the conditions of caregivers (Alemani *et al.*, 2016), shows that 19% of respondents on average have a university degree.
Moreover, formal education and training systems are not the only way in which people develop skills. This happens typically to people that need to start working at an early age and have no chance to acquire skills in a formal learning setting. Workers that own such uncertified skills may find more obstacle in compare to those who are at the same level of competences, but they have certified ones (Braňka, 2016).
Disadvantages groups (low-skilled, unemployed, low-income or migrant workers) would benefit from formal certifications but they lack financial resources to pay for them and have less awareness in the importance of skills recognition (Braňka, 2016).
Furthermore, "[L]a valutazione del sapere può essere affidata a diverse istanze. Sebbene la certificazione delle competenze sia un'attività istituzionale, da demandare a un organismo formalmente preposto, non tutte le competenze possono essere certificate anche se possono essere rilevate" (Meghnagi 2005, p. 109). For example, the tacit knowledge, such as the ability to learn or work in teams, which we have seen that is crucial in the era of automation, is more difficult to assess, codify and certify (Clark *et al.,* 2007).

Improving skills recognition comparability among countries and regions and developing a system that could update the systems automatically with real-time information (similar to the one proposed to deal with the problems related to the classification) could improve skills utilization, reduces skills mismatch and alleviates unemployment, poverty and inequality.

Appendix 3. – Polanyi's paradox: will it be overcome?

In the previous pages we focused on the reasons why the occupational polarization is less pronounced than one would have expected and why there are still, to say it with Autor, "so many jobs" (Autor 2014 and 2015)

This appendix returns specifically on Polanyi's Paradox with the aim to examine the much debated question whether the majority of tasks sooner or later will be automated. An understanding, although provisional and tentative, of this feature plays a significant role in comprehending the interplay between machine and human comparative advantages.

Autor explored two distinct paths that engineering and computer sciences are investigating in order to automate tasks: environmental control and machine learning. "The first path circumvents Polanyi's paradox by regularizing the environment, so that comparatively inflexible machines can function semi-autonomously. The second approach inverts Polanyi's paradox: rather than teach machines rules that we do not understand, engineers develop machines that attempt to infer tacit rules from context, abundant data, and applied statistics" (Autor 2015, p 23). Both lines of reasoning

are not able to fully accomplish nonroutine tasks and require certain forms of human-machine complementarity.

Most automated systems lack flexibility. Modern automobile plants – Autor emphasize – "employ industrial robots to install windshields on new vehicles as they move through the assembly line. But aftermarket windshield replacement companies employ technicians, not robots, to install replacement windshields. Evidently, the tasks of removing a broken windshield, preparing the windshield frame to accept a replacement, and fitting a replacement into that frame demand more real-time adaptability than any contemporary robot can cost-effectively approach" (p. 23).

The example he focuses on is that of the Google car.
"Computer scientists sometimes remark that the Google car does not drive on roads, but rather on maps. A Google car navigates through the road network primarily by comparing its real-time audio-visual sensor data against painstakingly hand-curated maps that specify the exact locations of all roads, signals, signage, and obstacles. The Google car adapts in real time to obstacles, such as cars, pedestrians, and road hazards, by braking, turning, and stopping. But if the car's software determines that the environment in which it is operating differs from the environment that has been preprocessed by its human engineers—when it encounters an unexpected detour or a crossing guard instead of a traffic signal—the car requires its human operator to take control. Thus, while the Google car appears outwardly to be adaptive and flexible, it is somewhat akin to a train running on invisible tracks. These examples highlight both the limitations of current technology to accomplish nonroutine tasks, and the capacity of human ingenuity to surmount some of these obstacles by re-engineering the environment in which work tasks are performed" (p. 24).

Up to present days Polanyi's paradox not allowing the programmer to build the appropriate computer instructions— "we know more than we can tell" has constituted an insurmountable barrier to automation, not allowing the programmer to build the appropriate computer instructions. Machine learning techniques can be a potential way out. "Where engineers are unable to program a machine to "simulate" a nonroutine task by following a scripted procedure, they may nevertheless be able to program a machine to master the task autonomously by studying successful examples of the task being carried out by others. Through a process of exposure, training, and reinforcement, machine learning algorithms may potentially infer how to accomplish tasks that have proved dauntingly challenging to codify with explicit procedures" (p. 25). The case taken into consideration is that of object recognition. "When training is complete, the machine can apply [a] statistical model to attempt to identify [objects] that are distinct from those in the original dataset. If the statistical model is sufficiently good, it may be able to recognize [objects] that are somewhat distinct from those in the original training data." (p. 25). Autor is open to several possible outcomes. "Some researchers expect that as computing power rises and training databases grow, the brute force machine learning approach will approach or exceed human capabilities. Others suspect that machine learning will only ever "get it right" on average, while missing many of the most important and informative exceptions." (p. 26).

The current research proposes the same question.
According to some technological optimists (Sussind, 2017), recent developments in machine learning have enabled the automatization of many non-routine tasks in a way that allow to overcome Polanyi's paradox.
They argue that machines are becoming able to make explicit more of the tacit knowledge and to learn tacit rules by themselves, by making use of statistics and learning from the context. According

to these researchers, therefore, computer systems don't need humans anymore. Scientists are indeed trying to use examples to teach machines to overcome Polanyi's paradox and have been successful to an extent. This view is supported by McAfee and Brynjolfsson (2016) who provides AlphaGo program, built by the Google subsidiary DeepMind, as an example of how the new approaches based on artificial intelligence are able to learn strategies entirely on their own, by seeing vast example of successes and failures of the game's matches, and to perform tasks based on tacit knowledge. In the 2016, AlphaGo program played against one of the world's top GO players, Lee Se-dol, and won.
Kothari (2017) concludes that "to overcome Polanyi's paradox, the deep neural networks employed to solve the paradox have become so complicated that this complexity has added uncertainty around how the machine arrives at an outcome creating another problem called 'The Black Box' problem". […] Scientists seek the answer to the black box in the fields of neuroscience and cognitive psychology. […] The answer to Polanyi's paradox may be learning through examples but the answer to the black box problem lies embedded within human tissue".

On the other hand, a reconfirmation of the relevance of the problem posed by Polanyi also derives from the observation of the current fields of application of artificial intelligence.

Ernst, Merola and Samaan (2018) that highlight that the three main groups of tasks that have become the focus of Artificial Intelligence (AI) applications are: matching tasks, classification tasks and process-management tasks. These three fields of applications of AI can be in turn categorized as (a) task-substitution; (b) task-complementarity; and (c) task expansion.
It is useful to describe in detail how the different groups are characterized and if and how the replacement of the work takes place in each of them.

- Matching tasks: are those that match supply and demand, especially on markets with a heterogeneous product and services structure in which machines have proved to be significantly faster and more efficient. Examples are ride-hailing services (Uber, Lyft), hotel and accommodation services (AirBnB, Booking.com) or human resource management (LinkedIn). Matching applications replace human activities with algorithms that allow to match supply and demand significantly faster and more efficiently.
- Classification tasks: are those applications of artificial intelligence concentrated on image and text recognition techniques. Some examples are facial recognition, partly in relation to the increase in surveillance cameras and techniques, medical applications (X-ray image diagnosing), legal services (reading and classifying legal documents), accounting and auditing (analysing balance sheets, fraud detecting), recruitment (screening applicants). Artificial intelligence-based applications for classification activities enable workers to focus on those requiring special attention, leaving to the computer the development of the most routine and repetitive ones.
- Process-management tasks: are a combination of the two previous sets of tasks, identifying patterns and bringing different suppliers and customers together along a supply chain, in combination with decentralized tracking and certification schemes ("blockchain"). There AI allows up-stream producers to integrate diversified supply chains through better information about product quality, certification schemes and market conditions. In this case, and AI-based applications often carry out tasks for which no human workforce was available to begin with, precisely because of the complexity of the tasks; in this case, the computer essentially expands the number of tasks that are

being carried out in an economy, thereby enhancing total factor productivity regardless of whether production is based mainly on skilled or unskilled labour.

Several questions still remain to be answered and more theoretical and empirical research needs to be developed to understand and forecast what machines are able or not to do. It is very likely, however, that the Polanyi paradox, in one form or another, will stay with us for a long time.

Appendix 4 – Industrial relations and changing job profiles: the disappearance of the job descriptions

Following the renewal of the CCNL for employees in the private metalworking and plant installation industries, changes have been made to the classification of workers. We believe that this change is an important step towards the awareness that technical progress is changing the way professions are defined: the labour market is in fact constantly changing and the figure of the worker assigned to mechanical and repetitive operations, which lend themselves to being classified in job descriptions, is disappearing in favour of highly qualified professionals capable of mastering complex technologies and systems[13]. The new CCNL no longer contains job descriptions, descriptions that in the '73 system introduced professional profiles (Negri and Pigni, 2015).

It can be seen that the renewal of the CCNL provides that each role can be classified into different levels, and different skills can be combined at the same level, as can be seen from the examples below[14] (Fig. 1):

"LEVEL D1:
Belongs to this level:
workers who perform elementary production, administrative or service activitiès related to a limited number of work positions of a specific operational/functional area according to defined work instructions. This role does not require specific professional knowledge and/or skills but does require basic digital, arithmetic and communication skills. Depending on the company contexts, these workers are coordinated in participating in company improvement initiatives.

LEVEL C1:
Belongs to this level:
workers with the characteristics of the previous level who carry out the activities of a specific work area of a specific functional operational field with versatility, recognised autonomy, with competence in specific technical diagnoses and in communication and teamwork. Depending on company contexts, they carry out activities of non-hierarchical operational liaison within the team or with related teams, tutoring and training alongside colleagues according to defined plans and procedures. They propose simple changes and adaptations and make an active contribution to improvement processes with autonomy in the application of the available methodologies".

[13] https://www.officeautomation.soiel.it/come-cambiano-le-professioni-dellindustria-4-0/

[14] https://studiosttr.it/wp-content/uploads/2021/06/Declaratorie-Dei-livelli.pdf

Attuali categorie	Minimi al 31/05/2021	Campi professionali	Nuovi livelli	Minimi al 01/06/2021
8ª	2.392,00	A Ruoli di gestione del cambiamento e innovazione	A1	2.424,86
7ª	2.336,02	B Ruoli specialistici e gestionali	B3	2.368,12
6ª	2.092,45		B2	2.121,20
5ªS	1.950,39		B1	1.977,19
5ª	1.819,64	C Ruoli tecnico specifici	C3	1.844,64
4ª	1.699,07		C2	1.722,41
3ªS	1.663,88		C1	1.686,74
3ª	1.628,69	D Ruoli operativi	D2	1.651,07
2ª	1.468,71		D1	1.488,89
1ª	1.330,54	Eliminazione 1ª categoria	-	-

Fig. 1: new workforce classification system. Source: https://studiosttr.it/news/la-nuova-classificazione-dei-lavoratori-nel-ccnl-industria-metalmeccanica-federmeccanica/

Figure 2 and the following example clearly show that the categorisation of levels is based on the level of autonomy, specific technical competence, polyvalence, multifunctionality, continuous improvement and innovation of the worker. We believe that this is a natural consequence of technical progress: the abandonment of the traditional job description is explained by the fact that roles can no longer be defined with respect to a set of tasks, but it is essential that they are defined with respect to a set of potential skills of the worker.

	D2	D3	C1
	Manutentore - ex "allievo"	Manutentore	Manutentore
Autonomia-responsabilità gerarchico-funzionale	NA	NA	Può svolgere ruoli di collegamento/affiancamento operativo senza responsabilità gerarchica di tipo "lear leader", "jolly" in alcuni modelli organizzativi
Competenza tecnico-specifica	Compiti elementari di pulizia tecnica, sostituzione guidata di componenti, semplici regolazioni guidate conoscenze generiche metodi manutentivi	Compiti ordinari nella specifica disciplina manutentiva di verifica, regolazione e sostituzione conoscenza sistemi manutenzione autonomi nella diagnostica nell'ambito di interventi ordinari intervenendo sulla base di procedure o strumenti, anche digitali, predefiniti.	Come per il D2 con riconosciuta autonomia
Competenze trasversali	Alfabetizzazione lingua italiana ed aritmetica Competenze digitali di base Orientamento su istruzioni lavoro in lingua	Limitata autonomia nella ricerca di semplici dati ed informazione tecniche con lettura dei principali formati di rappresentazione e nella compilazione di semplici rapporti di intervento preformattato anche attraverso strumenti digitali; alfabetizzazione nel glossario tecnico della lingua straniera rilevante;	Come per D2 con capacità di formazione per affiancamento sui colleghi
Polivalenza	Riferita ad uno specifico ambito di intervento	Su un limitato gruppo omogeneo di macchine - linee - sistemi	Sull'insieme di un gruppo omogeneo di apparti - macchine - linee - sistemi in funzione del contesto tecnologico
Polifunzionalità	Un solo ambito funzionale: es. meccanica, alimentazione elettriche, sensori	Eventuali elementi conoscitivi delle discipline manutentive complementari	Elementi conoscitivi base delle discipline manutentive complementari
Miglioramento continuo ed innovazione	Partecipazione guidata ad attività	Proposta di semplici modifiche ed adattamenti, partecipando in funzione del contesto azienda a gruppi di lavoro e miglioramento con l'utilizzo delle metodologie prescritte	Proposta di semplici modifiche ed adattamenti, partecipazione attiva a gruppi di lavoro e miglioramento con autonomia nell'utilizzo delle metodologie prescritte, ove previsti

Fig. 2: example of maintenance worker classification. Source: https://studiosttr.it/wp-content/uploads/2021/06/Esempio-inquadramento-manutentore-1.pdf

The following example shows the criteria for assessing the potential professionalism of the worker: "1.4. GLOSSARY AND EXAMPLES OF PROFESSIONAL CRITERIA

1. Hierarchical/functional autonomy/responsibility: this is the extent of impact and hierarchical or technical/functional influence and/or the degree of autonomy and executive discretion and initiative on the activities assigned in a given organisational context. This contribution may refer to

one's own individual activity or to other activities and resources (human, technological, economic, material, information, etc.).
Examples:
- Types of budget/accounting/outcome responsibilities etc.
- Size and qualification of reports in terms of breadth of functional hierarchical control
- Organisational area and complexity of the product/process on which influence is exercised (technical, economic, organisational and geographical)".

Appendix 5 – Covid-19 and the world of work

The advent of the pandemic emergency is having a considerable impact on the economy, society and daily lives. The consequences carried by this new phenomenon are significant for the economy and for the labour market, with substantial repercussions on the supply of goods and services, and on the consumer demand and investments. It is changing the destiny of lots of occupations, the skills' demand and the role of the technology and digitalization.
All enterprises are constantly facing severe challenges. Attention should be paid to some areas that can be identified as more vulnerable in enduring these circumstances, due to a considerable decrease in the overall turnover and an increase in job losses. Among them are accommodation services, catering, wholesale and retail, real-estate, and commercial activities. Considerably exposed are young people, women, service sector's employee, workers without a permanent contract (freelancers, self-employed, seasonal, and informal workers etc.) and immigrants (ILO 2020).
Nevertheless, there's been a fast increase in the demand of occupations necessary to deal with the current situation, for examples doctors, nurses and all the healthcare professions. Between the most in-demand jobs are also cleaning and sanitation, law enforcement and all occupations connected to the main needs of the community: transportation, agriculture, supermarket, pharmacies. Those people have been able to keep on practicing their activities but risk daily to contracting the COVID-19 (ILO 2020).
Some areas even gained profit in these circumstances: let's consider companies such as Amuchina (company producing disinfectant solutions for hands, surfaces etc.), Amazon, Netflix, pharmaceutical and biotechnology companies researching to develop a vaccine, telecommunication, enterprises producing equipment necessary for teleconferences, online educations and entertainment, profiles related to cybersecurity and risk management. All those saw an increase in the demand for their goods and services and have an economic benefit in this period from this situation.
Moreover, this situation has changed the approach to consider technology, openness to change and the way of working.
As was pointed out in the first chapters of this paper, the fourth industrial revolution was seen as a threat for many jobs and a cause of unemployment. In a context where the restrictive measures and the social distancing have been introduced to slow down the spreading of the virus, the way in which the technology is seen is changing. It is reconsidered as a support for citizens and occupation, as it is allowing employees to work from home. Smart working, telecommuting, videoconferences and e-learning were already being used before the advent of the COVID-19, but have now become essential to carry on the work activities. These evidences help people to look differently at the digital transformation and have learned to believe in the possibility and in the benefit of changes in the way they work.

The radical changes that already happened and that the advent of Covid-19 has increased, intensify even more the demand of digital and soft skills. Digital skills were becoming increasingly important, but were not yet taken into account by many companies (especially SMEs) that had not yet been digitized. Now their possession becomes fundamental no longer for a matter of competitive advantage, but for carry on digital work at a distance and thus becomes a matter of survival.
The changes in living habits and working spaces has increased the demand of adaptability skills such as time management, creativity, resilience, and anxiety management.

This situation also allowed to test digital work methods and identify their *pros* and *cons* in a concrete way: few aspects were overestimated and idealised by the workers, and many are the positive aspects that companies have been able to experience, which have allowed to break down some resistances that were glimpsed until just before. A lot of workers have hard time planning their daily schedule around their work duties, parting their private life from the work one, having to handle disturbances from relatives or flatmates. Of course, different personality traits reveal different perspectives. A more extroverted and outgoing person may suffer due to the lack of social dynamics present in the office, coffee breaks and talks with co-workers, and physical proximity to them. Others indeed appreciate the comforts of working from home and do not miss the office. They see this as an opportunity to slow down that hectic rhythm that led their daily life, and, by avoiding long commutes, benefit from the gained time for rediscovering personal hobbies, interests and passions.

Employers that usually preferred having a palpable control of the situation and a concrete overall view of the employees, found in working from home more advantages than they thought: decrease in bills and rent fees, and increase in the employee productivity and efficiency [10].

This situation of emergency gave a chance to companies to discover the pro and cons of smart working and look into the possibility of a future where working from home and from office, digital and traditional work, could be combined. Organisational structures will be re-evaluated as a consequence of the different opportunities arose and the distinct work methods tried to help a damaged work system in an attempt to recover during an emergency situation.

References

A ‘t Mannetje, H Kromhout (2003). “The use of occupation and industry classifications in general population studies.” International Journal of Epidemiology, Volume 32, Issue 3, Pages 419–428.

Acemoglu and Restrepo (2017). “Robots and Jobs: Evidence from US Labor Markets”. Working Paper 23285. National Bureau of Economic Research.

Acemoglu D. and Autor D. (2012). “What Does Human Capital Do? A Review of Goldin and Katz’s The Race between Education and Technology.” *Journal of Economic Literature,*50 (2): 426–63.

Acemoglu D. and Autor D.H. (2011). “Skills, Tasks and Technologies: Implications for Employment and Earnings” In *Handbook of Labor Economics*, Amsterdam: Elsevier. 4, part B:1043-1171.

Akst D. (2013), “What ca we learn from past anxiety over automation”, *The Wilson Quarterly*, Summer.

Alemani C., Marchetti S., Vianello F. (2016). “Viaggio nel lavoro di cura. Chi sono, cosa fanno e come vivono le badanti che lavorano nelle famiglie italiane”. Ediesse, Roma.

Alessi C., Barbera M., and Guaglianoni L., eds., (2019). *Impresa, lavoro e non lavoro nell’economia digitale*, Bari, Cacucci.

APCQ, (2017). *Process Classification Framework*. https://www.apqc.org/knowledge-base/documents/apqc-process-classification-framework-pcf-cross-industry-pdf-version-710.

Armaroli, P., Maranzana, P., Rinaldi, R., Salvioli, G., Vaccari, P. (2007). Il Sistema Regionale delle Qualifiche in Emilia-Romagna. Regione Emilia Romagna. http://www.form-azione.it/operatori/operatori.htm.

Arntz M., Gregory T. and Zierahn U. (2016). *The risk of automation for jobs in OECD countries: a comparative analysis.* OECD Social, Employment and Migration Working Paper 189.

Autor D.H. (2010). "The polarization of job opportunities in the US Labor Market: implications for employment and earnings", Center for American Progress and the Hamilton Project, May.

Autor D.H. (2014). "Polanyi's Paradox and the shape of employment growth", *NBER Working paper*, 20485, Cambridge (MA). http://www.nber.org/papers/w20485.

Autor D.H. (2015). "Why are there still so many jobs? The history and future of workplace automation", *Journal of Economic Perspectives*, 29(3): 3-30.

Autor D.H., Levy F., Murnane R.J. (2003). "The skill content of recent technological change: An empirical exploration". *The Quarterly Journal of Economics*, 118(4): 1279-1333.

Autor, D.-H. and Dorn D. (2013). *"*The growth of low-skill service jobs and the polarization of US labor market". *American economic Review,* 103(5), 1553-1557.

Azimov I. (1990). *Robot visions*, London, Victor Gollancz.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. Journal of management, Vol. 17, No. 1, 99-120.

Beblavý M., Akgüc M., Fabo B. & Lenaerts K. (2016). *What are the new occupations and the new skills? And how are they measured? State of the art report,* Working paper, Leuven, InGRID project, M21.6.

Beblavy M., Akguc M., Fabo B., Lenaerts K. (2016). "What are the new occupations and the new skills? And how are they measured?" European Commission, InGRID (Inclusive Growth Research Infrastructure Diffusion) Working paper.

Becattini G. and Rullani E. (1993) "Sistema locale e mercato globale", Economia e politica industriale, n. 80, pp. 25-49.

Becattini G. and Rullani E. (1993) "Sistema locale e mercato globale", *Economia e politica industriale*, n. 80, pp. 25-49.

Becattini G. and Rullani E. (1993). *Sistema locale e mercato globale*, *Economia e politica industriale*, n. 80, pp. 25-49.

Belussi, F. (1999). Policies for the development of knowledge-intensive local production systems. Cambridge Journal of Economics, 23, 729-747.

Benadusi L. and Molina S., eds, (2018). *Le competenze. Una mappa per orientarsi*. Fondazione Agnelli. Bologna, il Mulino.

Berger T. and Frey C.B. (2016). *Digitalisation, jobs and convergence in Europe. Strategies for closing the gap*, Prepared for the European Commission, Oxford Martin School, University of Oxford.

Berman E., Bound J., Machin S. (1998). Implications of Skill-Biased Technological Change: International Evidence, The Quarterly Journal of Economics, Volume 113, Issue 4, Pages 1245–1279.

Bessen, James E., Industry Concentration and Information Technology (2017). Boston Univ. School of Law, Law and Economics Research Paper No. 17-41.

Bowles J. (2014). *The computerization of European jobs*, Technical report, The Bruegel Institute.

Braňka, Jiří (2016). "Understanding the Potential Impact of Skills Recognition Systems on Labour Markets: Research Report". International Labour Office, Skills and Employability Branch, Geneva.

Brusco S. (1994), "Sistemi globali e sistemi locali", Economia e politica industriale, n. 84, pp. 63-76.

Brusco S. (1994). *Sistemi globali e sistemi locali*, Economia e politica industriale, n. 84, pp. 63-76.

Business Insider Italia (2017). *"Il 65% dei bambini che iniziano le elementari farà un lavoro che oggi non esiste. E allora, che cosa deve insegnare la scuola oggi?"* https://it.businessinsider.com/il-65-dei-bambini-iniziano-le-elementari-fara-un-lavoro-che-oggi-non-esiste-e-allora-che-cosa-deve-insegnare-la-scuola-oggi/?refresh_ce.

Campagna L., Pero L. e Ponzellini A.M (2017). Le leve della innovazione. Lean, partecipazione e smart working nell'era 4.0, Guerini Next (2019).

Campagna L., Lizza M. e Pero L. (2019). La fabbrica delle competenze e della dignità. Idee e progetti per il PNRR: Next Generation Italia, Edizioni Lavoro.

Camuffo, A., Grandinetti, R. (2011). I distretti industriali come sistemi locali di innovazione. Sinergie rivista di studi e ricerche, 24(69): 33-60.

Cattani, L., Purcell, K. & Elias, P. (2014). *SOC(HE)-Italy: A Classification for Graduate Occupations.* SSRN Electronic Journal.

Cedefop (2006). ICT skills certification in Europe. Cedefop Dossier Series, 13. Luxembourg: Office for Official Publications of the European Communities.

Cedefop (2009). *European guidelines for validating non-formal and informal learning*. Luxembourg: Office for Official Publications of the European Communities.

Cedefop (2013). *Quantifying skill needs in Europe. Occupational skills profiles: methodology and application*. Luxemburg: Pubblications Office of the European Union.

Cedefop (2014). Use of validation by enterprises for human resource and career development purposes. Luxemburg: Pubblications Office of the European Union.

Chiacchio F., Petropoulos G., Pichler D. (2018), "The impact of industrial robots on EU employment and wages: a local labour market approach", Bruegel Institute, Working Paper Series, n. 2.

Cipriani A., Gramolati A. and Mari G., eds., (2018). *Il lavoro 4.0. La quarta rivoluzione industriale e le trasformazioni delle attività lavorative*. Firenze, Firenze University Press.

Clark, P F, J B Stewart, e D A Clark (2007). "Portability of Skills". International Labour Office, Governing Body. Committee on Employment and Social Policy, Geneva.

Coase R.H. (1937). "The Nature of the firm". *Economica*, november, pp. 386-405.

Coase R.H. (1988). "The Nature of the firm: Origin". Journal of Law, Economics and Organization, vol. 4, n. 1, pp. 3-17.

Cominu, S. (2018). "Tutti knowledge worker? Ricchezza e impoverimento dei lavori". *Sociologia del lavoro*, 151:174-189.

Crockett T, Lin E., Gee M., Sung C. (2018). *Skills-ML: An Open Source Python Library for Developing and Analyzing Skills and Competencies from Unstructured Text*. Center for Data Science and Public Policy, The University of Chicago Press.

Crosby O. (2002). *New and Emerging Occupations*, Occupational Outlook Quarterly, Fall 2002: 17-25.

E. Ernst, R. Merola, D. Samaan (2018). "The economics of artificial intelligence: Implications for the future of work". ILO Future of Work research paper series.

Ernst E., Merola R. and Samaan D. (2018). "The economics of artificial intelligence: Implications for the future of work". *ILO future of work research paper series*.

European Commission (2013). *ESCO – European Classification of Skills/Competences, Qualifications and Occupations*. Luxembourg: Publications Office of the European Union.

European Union (2019). *ESCO handbook*. Luxemburg: Pubblications Office of the European Union.

Ferrari L., de Laurentis D. and Scuteri A. (2019). La consulenza finanziaria 3.0. Dinamiche relazionali e tecniche di gestione alla luce della nuova finanza comportamentale, Alessandria, Vicolo del Pavone.

Ferrari, F. (2015). Innovazione tecnologica nei distretti industriali e nei cluster tecnologici: analisi dello sharing di knowledge nel distretto ceramico di Modena e Reggio Emilia. LUISS, Dipartimento di Impresa e Management.

Fontana D., Paba S. and Solinas G (2019), "Working conditions and quality of work in the digitized factory" in Marcuzzo MC., Palumbo A. and Villa P. (eds), Economic Policy, Crisis and Innovation: Beyond Austerity in Europe, London, Routledge, 2019, pp. 219-232. (ISBN: 9780367260293).

Franzosi, C., Mandrone, E, Premutico, D., Pitoni, I., Angotti, R., Carlini, A., Daniele, L., Penner, F., Premutico, D.; Scalmato, V., Spigola, C., Vaccaro, S. Mereu, M.G., Mazzarella, R. (2016). Istruzione e formazione in Europa, Italia. Rapporto pleliminare. A cura di Cedefop e INAAP.

Freddi D. (2017). *Gli effetti occupazionali della digitalizzazione – Rassegna della letteratura*. Ricerca europea Transform Europa and Rosa Luxemburg Stiftung 2016-2017.

Fretwell D. H., Morgan V. L., Arjen D. (2001). *A framework for Defining and Assessing Occupational and Training Stardards in Developing Countries.* Information Series No. 386. World Bank, The Ohio State University, European Training Foundation.

Freud S. (1913). "Inizio del trattamento", in. Totem e tabù e altri scritti, Opere, vol. VII, trad. it. Torino, Bollati Boringhieri, 1989.

Frey C.B. and Osborne M.A. (2013). "The future of employment: How susceptible are jobs to computerization?" Oxford Martin Programme on The Impact of Future Technologies, University of Oxford. Revised in *Technological Forecasting and Social Change,* 2017, 114: 254-280.

Gallino L. (1993), Dizionario di sociologia, Milano, Utet.

Garibaldo F. e Rinaldini M., a cura di, (2021), Il lavoro operaio digitalizzato. Inchiesta nell'industria metalmeccanica bolognese, Bologna, il Mulino.

Gatti M., Garbellini N. and Garibaldo F. (2018), *Industry 4.0 and its consequences for work and labour*. Roma: Fondazione Claudio Sabattini.

Gertler, M. (2003). Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there). Journal of Economic Geography, 3(1), 75-99. Retrieved May 15, 2020, from www.jstor.org/stable/26160465.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there). Journal of Economic Geography, 3(1), 75–99.

Ghergo, F. (2011). Storia della formazione professionale in Italia 1947-1997. Volume II. CNOSFAP e Ministero del Lavoro e delle Politiche Sociali.

Graetz G, Michaels G (2018). Robots at work. The Review of Economics and Statistics. 100(5): 753–768.

Grant K. A. (2007) "Tacit Knowledge Revisited – We Can Still Learn from Polanyi" The Electronic Journal of Knowledge Management Volume 5 Issue 2, pp 173 - 180, available online at www.ejkm.com.

Gregory, T., Salomons A. and Zierahn U. (2016). Racing with or against the machine? Evidence from Europe. *ZEW Discussion Paper*, 16-053. http://ftp.zew.de/pub/zew-docs/dp/dp16053.pdf.

Hernandez Ospino G. C. (2018). *Occupations: Labor Markek Classifications, Taxonomies, and Ontologies in the 21st Century.* Technical note n° (IDB-TN-1513). Labor Markets Division.

Hernandez, Carlos G. Ospino (2018). "Occupations: Labor Market Classifications, Taxonomies, and Ontologies in the 21st Century". Cataloging-in-Publication data provided by the Inter-American Development Bank. Felipe Herrera Library, Ospino, Carlos.

Hoffmann E. (1999). *International Statistical Comparison of Occupational and Social Structures: Problems, Possibilities and the Role of ISCO-88*. http:// www.ilo.org/public/english/bureau/stat/papers/index.htm.

Hoffmann E. and Scott M. (1993). *The revised international standard classification of occupation*. Geneva, International Labour Office. Bureau of Statistics.

https://en.wikipedia.org/wiki/Tacit_knowledge

ILO (2015). *'Introduction to Occupational Classifications'* International Standard Classification of Occupations (http://www.ilo.org/public/english/bureau/stat/isco/intro.htm)

ILO (2020). COVID-19 and the world of work: Impact and policy responses. ILO Monitor 1st Edition.

ILO (2020). COVID-19 and the world of work. Updated estimates and analysis. ILO Monitor 2nd Edition.

International Standard Classification of Occupations: ISCO-08 (2012). International Labour Office, Geneva.

Isfol (2015). Validazione degli apprendimenti: standard, processi, garanzie nelle riforme nazionali. Roma.

Istat (2013). *La classificazione delle Professioni*. A cura di F. Gallo e P. Scalisi.

Istat (2016). La classificazione delle Professioni.

Jin D. J. And Stough R. R. (1998). *Learning and learning capability in the Fordist and post-Fordist age: an integrative framework*. Environment and Planning A, 30, 1255-1278.

Jon Kabat-Zinn (2005). "Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life", Hyperion.

Katz, M. B. (1972). "Occupational Classification in History." *The Journal of Interdisciplinary History* 3, no. 1: 63-88.

Kiley, M. T. (1999). The Supply of Skilled Labour and Skill-biased Technological Progress. The Economic Journal, 109(458), 708–724.

Kilhoffer Z. (2020). *Report on how to identify and compare newly emerging occupations and their skill requirements*, Deliverable 12.2, Leuven, InGRID-2 project 730998 – H2020.

Kothari (2017). Polanyi & The Black Box. Clues To The Future of AI & Mankind. https://medium.com/@abhishekkothari/polanyis-paradox-b197b61a85b1.

Koucký J., Kovařovic J., Lepič M. (2012). *Occupational Skills Profiles: Methodology and application. Contribution to the concept, structure, and quantification of skill needs in Europe*. Education Policy Centre (EPC), Charles University in Prague.

Levy, F. and Murnane, R. (2004). The new division of Labor: How computers are creating theNext Job Market. Princeton University Press.

Lodigiani R. (2011). *Il mito delle competenze tra Procuste e Prometeo*. Quaderni di Sociologia, LV(55), 139-159.

Lodigiani, Rosangela. Sarli, Annavittoria. (2017) Migrants' competence recognition systems: controversial links between social inclusion aims and unexpected discrimination effects. *European Journal for Research on the Education and Learning of Adults*, 8.

London Electoral History – Steps Towards Democracy. Classification by Occupation. http://leh.ncl.ac.uk/PDF%27s/LEH-Classification/LEH-CLASSIFICATION7.11OCCUPATIONS.pdf.

M. Franceschetti, D. Guarascio, M. Grazia Mereu (2019). "Fabbisogni professionali e competenze per il lavoro che cambia. L'indagine pec-inapp su professioni e competenze nelle imprese". INAAP Policy Brief.

Magnaghi, G. (2020). Come cambiano le professioni nell'industria 4.0. https://www.officeautomation.soiel.it/come-cambiano-le-professioni-dellindustria-4-0/

Manyika J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., Dewhurst, M. (2017). A future that works: automation, employment, and productivity. McKinsey Global Institute. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Full-report.ashx.

Martin, M. E. (1967). *Occupational Classification*. Demography 4, 843–845.

Maskell, P. & Malmberg, A. (1999). *Localized Learning and Industrial Competitiveness*. Cambridge Journal of Economics. 23. 167-85.

Mazzarella F., Mallardi F., Porcelli R. (2017), Atlante lavoro. Un modello a supporto delle politiche dell'occupazione e dell'apprendimento permanente, Sinappsi, 7, n. 2-3, pp. 7-26.

Mazzarella R. and Porcelli R. (2017). *Procedura per la manutenzione (aggiornamento e sviluppo) dell'Atlante del lavoro e delle qualificazioni con riferimento al Decreto interministeriale del 30 giugno 2015*. Roma, INAPP.

Mazzarella, R., Porcelli, R. (2017). Procedura per la manutenzione (aggiornamento e sviluppo) dell'Atlante del lavoro e delle qualificazioni con riferimento al Decreto interministeriale del 30 giugno 2015. INAAP.

McAfee A. and Brynjolfsson E. (2016). "Where Computers Defeat Humans, and Where They Can't". *The New York Times*, 16 march, Retrieved, 2018-10-04.

Meghnagi S. (2005) Il sapere professionale. Competenze, diritti, democrazia, Milano, Feltrinelli editore.

Meghnagi S. (2005). *Il sapere professionale. Competenze, diritti, democrazia*, Milano, Feltrinelli editore.

Meghnagi S. and Mora F. (2018), "Competenze nel mondo del lavoro", in Benadusi L. and Molina S. (2018), pp. 63-83.

Mineo S. and Piperno I., eds, (2015). *PIAAC – Formazione e competenze online*. Roma, ISFOL. http://www.oecd.org/skills/ESonline-assessment/abouteducationskillsonline/EducationSkillsOnline%20Info_italian_version%20(3).pdf.

Mokyr J., Vickers C. and Ziebarth N. L. (2015). "The history of technological anxiety and the future of economic growth: Is this time different?", *Journal of Economic Perspectives*, 29(3): 31-50.

Moravec H.P. (1988). *Mind children: the future of robot and human intelligence*, Cambridge, Harvard University Press.

Moretti E. (2014), *La nuova geografia del lavoro*, Milano, Mondadori, (US ed., *The new geography of jobs*, 2012).

Moskowitz R. and Chief B. (2017). A Brief History of Occupational Classification in the United States, Part 2. Bureau Chief.

National Research Council (1980). *Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles*. Washington, DC: The National Academies Press.

Nedelkoska L. and Quintini G. (2018) Automation, skills use and training. In OECD Social, Employment and Migration Working Papers, No. 202, OECD Publishing.

Negri, S. and Pigni, G. (2015). Il nuovo sistema di inquadramento e classificazione dei lavoratori nell'ipotesi di rinnovo del 5 febbraio 2021. Bollettino speciale ADAPT, n. 1.

OECD (2012), Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies, OECD Publishing.

Orpinas P. (2010). Social Competence. 10.1002/9780470479216.corpsy0887.

Ospino Hernandez C.G. (2018). *Occupations: Labor Market Classifications, Taxonomies, and Ontologies in the 21st Century.* Cataloging-in-Publication data provided by the Inter-American Development Bank. Felipe Herrera Library. (IDB Technical Note; 1513).

Paba S. and Solinas G. (2018). "In favour of machines (but not forgetting the workers)". In Ales E., Curzi Y, Fabbri T., Rymkevith O., Senatori I. and Solinas G., eds, *Working in Digital and Smart Organizations – Legal, Economic and Organizational Perspectives on the Digitalization of Labour Reations*, London: Palgrave-Macmillan.

Parsons C., Rojon S., Samanani F., & Wettach L. (2014). *Conceptualising International High-Skilled Migration*. IMI Working Papers Series, No. 104.

Pero L. (2019). "Il lavoro al tempo dell'industria 4.0", Milano, Scuola di cultura Politica, https://www.youtube.com/watch?v=ahE3Y144K6E

Pero L. e Campagna L. (2020). Innovazione e Impresa 4.0, Cento Studi Cisl, https://www.youtube.com/watch?v=FnHfJ77NX_A

Perulli E. (2013), ed., *Validazione delle competenze da esperienza: approcci e pratiche in Italia e in Europa*, Roma, Collana Isfol *I Libri del FSE*.

Polanyi M. (1966), *The Tacit Dimension*, University of Chicago Press, Chicago.

Polanyi, M. (1958). Personal Knowledge: "Towards a Post-Critical Philosophy. University of Chicago Press, Chicago.

Rumelt, R. P. (1991). How much does industry matter? Strategic Management Journal, 12(3), 167–185.

Sgobbi F. (2019). "La polarizzazione del lavoro nell'era digitale: un'analisi empirica del caso italiano", in Alessi C. e altri, pp. 251-273.

Siekmann G., & Fowler C. (2017). *Identifying work skills: international approaches*, NCVER, Adelaide.

Simoncini G.R, (2016). "Il modello americano dell'Occupational Information Network e dell'Occupational Licensing", Roma, ADAPT.

Stinchcombe, A. L. (1990). *Information and Organizations*. Berkeley: University of California Press.

STTR Tax & Labour S.A.S. (2021). La nuova classificazione dei lavoratori nel CCNL industria metalmeccanica federmeccanica. https://studiosttr.it/news/la-nuova-classificazione-dei-lavoratori-nel-ccnl-industria-metalmeccanica-federmeccanica/

Susskind, Daniel (2017), "Re-Thinking the Capabilities of Machines in Economics", University of Oxford Department of Economics Discussion Paper Series 15127 , Oxford.

Tecnostruttura (2015). Il lavoro del Gruppo Tecnico competenze nell'ambito della strategia nazionale per la costruzione del sistema nazionale di certificazione delle competenze. Pubblicazione on line della Collana ADAPT.

TextySrl and University of Pisa. "Smart(er) or Hard(er) work per i knowledge workers?". Survey results.

Van Leeuwen H. D. M., Maas I. & Miles A. (2004). *Creating a Historical International Standard Classification of Occupations An Exercise in Multinational Interdisciplinary Cooperation*, Historical Methods: A Journal of Quantitative and Interdisciplinary History, 37:4, 186-197.

Van Leeuwen, Marco H. D., Ineke Maas, and Andrew Miles (2004). "Creating a Historical International Standard Classification of Occupations An Exercise in Multinational Interdisciplinary Cooperation." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 37, no. 4. 186–97.

Van Vulpen E. (2020). *Job Classification: A Practicioner's Guide.* AIHR Academy.

WEF – World Economic Forum (2018), *The future of work report 2018*, Cogogny (Geneva): Centre for the New Economy and Society.

Sitography

[1] http://career.iresearchnet.com/career-assessment/occupational-classification-systems/.
[2]http://www.integrazionemigranti.gov.it/normativa/Pagine/Atlante-del-lavoro-e-Repertorio-Nazionale.aspx.
[3] https://www.onetonline.org.
[4] https://ec.europa.eu/esco/portal.
[5] https://atlantelavoro.inapp.org/repertorio_nazionale_qualificazioni.php.
[6] http://escobadges.eu.
[7] https://openbadges.org.
[8] https://woman.thenest.com/improve-personal-competencies-9345.htm.
[9] https://it.businessinsider.com/il-65-dei-bambini-iniziano-le-elementari-fara-un-lavoro-che-oggi-non-esiste-e-allora-che-cosa-deve-insegnare-la-scuola-oggi/?refresh_ce.
[10] https://www.corriere.it/dataroom-milena-gabanelli/coronavirus-smartworking-connessione-oltre-11-milioni-italiani-senza/deb45d24-66e8-11ea-a26c-9a66211caeee-va.shtml.
[11] https://www.mckinsey.com/industries/public-sector/our-insights/how-to-rebuild-and-reimagine-jobs-amid-the-coronavirus-crisis.

[12] http://leh.ncl.ac.uk/PDF%27s/LEH-Classification/LEH-CLASSIFICATION7.11OCCUPATIONS.pdf.
[13] http://formazionelavoro.regione.emilia-romagna.it.
[14] http://www.integrazionemigranti.gov.it/normativa/Pagine/Atlante-del-lavoro-e-Repertorio-Nazionale.aspx.

Personal Communications

D. Teloni (Managing Director), interview with authors, Feb. 4, 2019.
A. Papantoniou (Cognizone CEO), interview with authors, Mar. 5, 8, 2019.
R. Trainito (Senior Manager, PwC), interview with authors, Jul. 12, 2019.
G. Fantoni (Associate Professor at University of Pisa), interview with authors, Sept. 9, Nov. 27, 2019.
F. Bergamini (Director at Emilia-Romagna Region), interview with authors, Feb. 13, 2020.
D. Venturelli (Scientist at NASA, Founder of Archon), inspiration from the interview Heroes On Air, meet in Maratea (https://www.facebook.com/watch/live/?v=1146357295496504&ref=watch_permalink).

From HRM to HRM 4.0: a systematic literature review of main topics and values behind HR analytics

Abstract

In the human resource (HR) management literature, over the past three decades, a shared consensus has developed that people are the most valuable asset for an organization's investment and the greatest source of competitive advantage. Despite this agreement, HR professionals had always difficulties in demonstrating the value of their function and in being recognized as a key contributor. We present a systematic review of 91 studies in which we analyze the development of HR analytics and workforce measurement research over time and identify important trends, the major topics and themes and progresses over time, the purpose and applications of HR metrics and analytics and how do them produce value for the organization. Furthermore, we propose a typology contributed to the understanding of the areas of focus in the literature.
Our findings suggest that high performance HR systems and HR analytics practices are a powerful way to help the HR function to explain the value of its profession, to support data-driven decision-making, strategic workforce management and planning and management practices, to make a positive changes for the workforce, and thus to produce a value addition and a competitive advantage for the company. Much of the research to date does not investigate the interrelationships between the dimensions that characterize the literature and some of the major areas. Overall, we thus still know little about some of the topics and how synergies and interaction between the various domains operate. We offer actionable suggestions regarding the current state of the HR analytics and measurement literature and the areas that are still to be explored.

Introduction

There is a growing body of literature that recognises the importance of people as the most valuable asset for an organization's investment (Srimannarayana, 2010; Boudreau, 1998; Du Plessis and De Wet Fourie, 2016; Nienaber and Sewdass, 2016; Chattopadhyay *et al*., 2017) and the greatest source of competitive advantage (Tootell *et al*., 2009; De Mauro *et al*., 2018; Boudreau, 1998). However, HR professionals had always difficulties in demonstrating the value of their function (Kryscynski *et al*., 2017; Pilenzo, 2009). One explanation is that HR fall short of being a strategic partner because it lacks the types of analytics and data-based decision-making capability to influence business strategy (Kryscynski *et al*., 2017). Recently, the continuous digitalization of the relationship models between company and employees makes an increasing number of information available to the Human Resources function, which can be enhanced through new tools and methodologies to improve the attractiveness, evaluation and development of human capital. This growing trend is demonstrated by data on the number of publications over time (Figure 1), that shows an exponential increase from 2008. Today's HR transformation is a direct result of the rise of the fourth industrial revolution, that has led to a radical digital transformation of companies that increasingly produces a change in business processes, which become able to take advantage of the intelligence introduced by digital technologies in most of the company's activities.

The growth in the number of information sources available and the relative amount of data produced, together with the availability of more powerful and affordable processing and storage technology, has brought awareness in large companies of how Big Data analysis can represent a source of competitive advantage and a tool of evolution of the same business model. The data-

driven culture allows the HR departments to make more rational, productive and strategic choice. In light of the rapid changes in technology and the environment, the conventional HR metrics are only useful for answering simple questions (Handa and Garima, 2014).
As might not be surprising in a nascent discipline, numerous terms and synonyms are used to describe HR Analytics. In addition to HR Analytics, the terms HR Metrics, HR Analytics, Talent Analytics, Human Capital Analytics, and People Analytics have all been used to describe this field (Huselid, 2018). One of the most cited definitions is that of Marler and Boudreau (2017: 15), which define HR analytics as: "A HR practice enabled by information technology that uses descriptive, visual, and statistical analyses of data related to HR processes, human capital, organizational performance, and external economic benchmarks to establish business impact and enable data-driven decision-making".

HR analytics has several goals and applications (Chalutz, 2019). While some research has been carried out on HR metrics or analytics and on the various goals and applications, no studies have been found which investigate how the literature and its focus has evolved over time.
There are two primary aims of this study. The first is to provide an integrative analysis of the literature both on the conventional HR metrics and on the modern sophisticated analytics techniques; the second is to identify where new research is needed. This work aims to aid both researchers and practitioners with respect to the uses, applications and benefits of the HR analytics techniques.
To this end, our paper asks and answer three research questions:

RQ1. What are the major topics and themes that have been developed within HR analytics and workforce measurement research and how they have progressed over time?

RQ2. What are the purposes and applications of HR metrics and analytics and how do them produce value for the organization?

RQ3. Which topics are already much explored and on which, on the other hand, should future studies be focused?

This article has the following structure. The next methodological section outlines the systematic literature review approaches and steps. Then, the results part presents the findings of the research, focusing on the key themes that topics that emerged through the examination of the literature, answering the first and the second questions. Finally, in the discussion and implication section we answer the last research question by proposing a typology to classify the literature and identifying which areas are over explored and which, on the other hand, need to be more investigated.

Method

We drew on the definition of Green and Higgins (2005, cited in Moher *et al*., 2009, p. 1), which defined the systematic review as "a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review". The important advantages of using this approach is that it provides a replicable, scientific and transparent process that allow to conduct a comprehensive, unbiased search (Tranfield *et al.*, 2003). We prepared the review and the synthesis of the literature according to the procedure recommended by Tranfield *et al.* (2003) that consists of three main stages: planning, executing, and reporting, as outlined below.

Planning

In accordance with the process of the systematic literature review, the first step was to identifying the key data sources that are consistent with the research's purpose. We considered only scholarly peer-reviewed journal articles written in English. Since the objective of this study was to develop an integrated framework of People Analytics, which is an interdisciplinary field, we searched articles from two different type of databases: the first belongs to the Human Resource Management domain and the second belongs to the Information Science domain. Commenting on research in this area, Tursunbayeva, Di Lauro and Pagliari (2018, p. 230) observe: "While most articles come from the Business, Management and Accounting domains, social science has remained prominent, and the importance of the computing and data sciences is increasingly evident, echoing the growth of HRIS and digital innovations for monitoring, evaluating and predicting work". We ran our search in the electronic catalogue OneClick and from each domain, we selected the relevant associated categories in Business, Economics and Economic and Business Sciences from the first one and Computer Science, Statistics and Engineering from the second one. OneClick[15] is the Discovery Tool for the integrated search from subscribed full text-databases, citation databases, institutional repositories and the library catalogue of the University of Modena e Reggio Emilia. The databases that resulted through our search were Business Source Ultimate (EBSCO), Econlit with full text (EBSCO) (Business), Emerald Insight, Oxford University Press Journals (Economics), Business Source Ultimate (EBSCO) (Economic and Business Sciences), ACM (Association for Computing Machinery), MathSciNet (via EBSCOhost) (Computer Science and Statistics), Elsevier Journal ScienceDirect, IEEE Xplore, Scopus and Web of Science (Engineering).
We reviewed all articles independently to determine whether they met our predefined criteria, which are illustrated in the following paragraph, and then discussed ambiguous cases to achieve agreement.

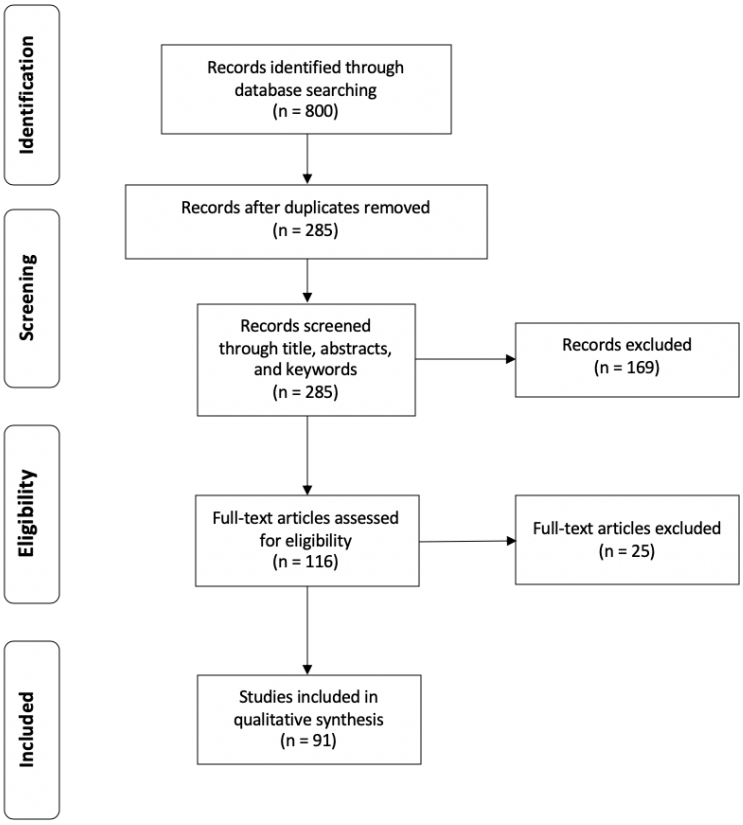
Executing

To guarantee transparency, we base the selection of relevant contributions on the specific inclusion criteria established. The first step in this process was to determine the keywords relevant to our study. To identify these keywords, we started from a list of studies about People Analytics and HR Analytics, we selected the articles published in scholarly peer-reviewed journals in English and we create a list of the articles' key terms. We used them as a starting point to create our final list of terms: *People Analytics*, *HR Analytics*, *Workforce Analytics*, *Analytics in HR*, *HR Metrics*, *HR Measurement*, *Human Capital Analytics*, *Human Resource Metrics*, *Predictive Analytics for Human Resource*, *Workforce Metrics*, *Big Data* "AND" *People*, *Big Data* "AND" *HR*. For each database, we searched the title and abstract with the selected terms combined using the Boolean "OR" operator. This primary search produced more than 1819 articles (Business Source Ultimate: 358, Econlit with full text: 21, Emerald Insight: 85, Oxford University Press Journals: 10, MathSciNet: 1, ACM: 37, Elsevier Journal ScienceDirect: 396, IEEE Xplore: 66, Scopus: 805, Web of Science: 41). The primary list of articles was refined by eliminating articles that did not meet the inclusion criteria, especially in those databases where it was not possible to filter the search (e.g., to impose the scholarly peer-reviewed restriction). Through this manual process, we removed 1019 articles from the initial set. Most of the excluded articles where conference proceedings or articles where HR was used as an abbreviation for something else (e.g. heart rate).

[15] http://www.libraries.unimore.it/site/home/research-tools/discovery-tool-oneclick.html

Next, we used the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram for selecting the articles to include in the review (Figure 1).
We began the selection process by removing duplicate data. Following this, we reviewed the article title, abstracts, and keywords to select relevant studies. Finally, if the paper still appeared relevant, we read the whole publication to determine whether the article was suitable for inclusion in the literature review. This final filter left 91 articles that fully met the inclusion criteria.

Figure 1
Prisma flow chart visualising the article selection process



Reporting

In this phase, we produce a synthetic description of the studies and then we report the results of the analysis.
More than 80% of the articles included in the analysis were published after 2012, as can be seen from the figure below, indicating that the topic of HR/People Analytics and workforce Measurement is relatively new.

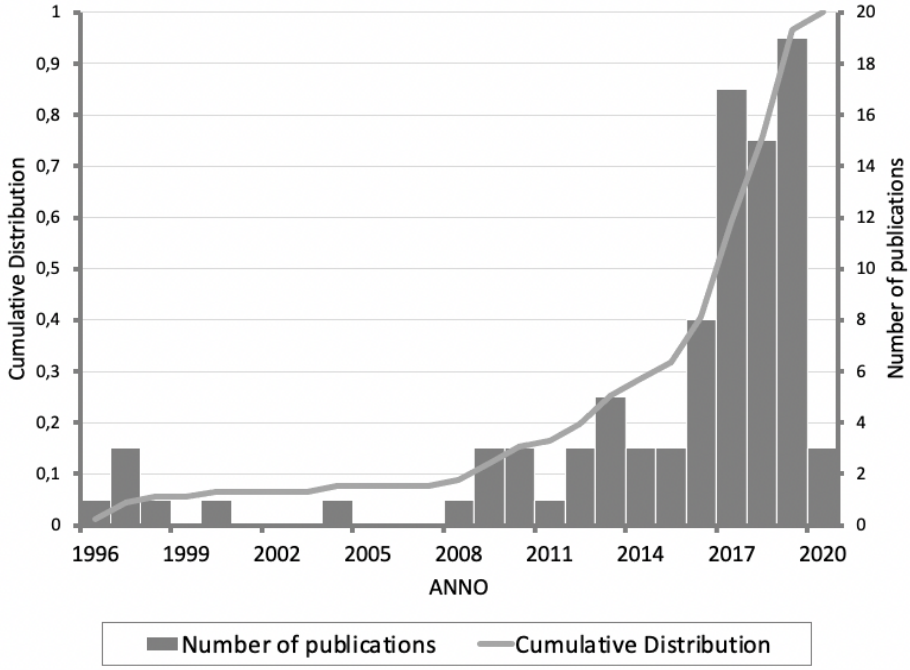


Figure 1: Time trend of publications on HR Analytics and workforce measurement from 1996 to 2020 (n=91)

We classified the journal of publication into three disciplinary areas: HRM, management and business and engineering.

Nearly half of the articles pertain to the HRM area, just over a third to the management and business domain, more than 5% to the engineering field and around 16% to other areas, such as operational research or accounting and finance (see Figure 2).

The articles were published in fifty-five journals, and the three journals that contained the large part of the articles were: *Human Resource Management* (n=12), *Journal of Organizational Effectiveness: People and Performance* (n=5), and *Human Resource Management Journal* (n=4).

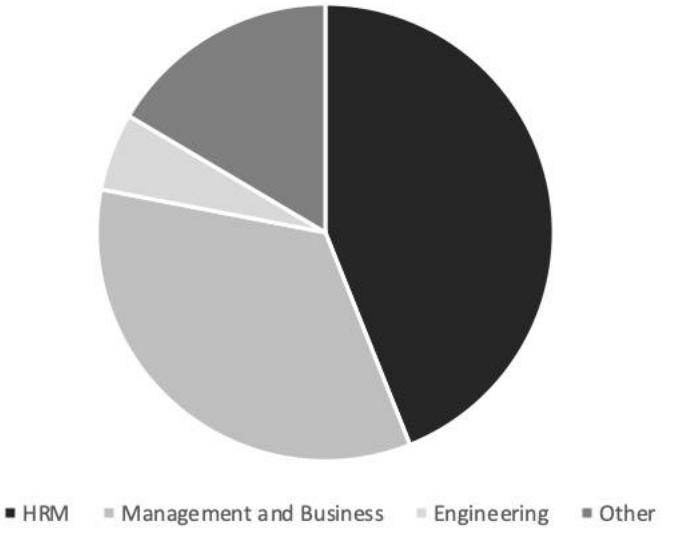


Figure 2: Frequency distribution of articles by disciplinary area: Articles aggregated by journal and journals aggregated by field (n =91)

Only 29% (n=26) of the study were conducted in a specific country. Most of them were conducted in Europe (n=7) and Asia (n=8), but some were conducted in North America (n=5), Oceania (n=4), and Africa (n=2).
Most of the research articles are conceptual articles (45%) and empirical articles (41%). Only a minority of the research articles are case based (10%) or technical (4%).
Most of the studies' data were collected through questionnaire and surveys (n= 21) of HR managers and professionals, while others were collected through documents and records (n=8), interviews (n=7), observations (n=1), focus groups (n=1) and a combination of the previous method (n=6).

Emergence and evolution of HR analytics research

We adopted a part of the classification used by Chalutz (2019) to investigate the evolution of the HR analytics research. The results of our research, displayed in Figure 1 and Table 1, indicates that the development of the literature of HR analytics and workforce measurement followed three different periods and increased over time. The first article in our sample of literature was published in 1996. The first is a period of incubation (1996-2000) during which there are just a few articles per year and almost 8 percent of the HR analytics and workforce measurement research was published. The second was stationary period (2001-2007), during which only one article was published. Then finally, starting from 2008, there is a period of exponential growth that indicated an increasing interest in HR analytics, when 92 percent of HR analytics and workforce measurement research was published. We have divided this period into two further periods: a period of incremental growth (2008-2016), during which almost 33 percent of the articles was published, and a period of substantial growth (2017-2020), that included 59 percent of the articles. The 2020 counts few articles because we finished our data collection at the beginning of the year.
The results and the development trajectory are consistent with the previous research and confirms the growing interest in the field of HR analytics (Chalutz, 2019; Marler and Boudreau, 2017). Moreover, the focus of the HR analytics and workforce measurement research has changed over time. While early publications examined HR metrics and workforce measurement from a narrow economic and personnel perspective (Chalutz, 2019), in recent years the research evolved into sophisticated HR analytics techniques and practices that are able to influence not only the HR function, but also the organizational and business outcomes, which has transformed it into a multidisciplinary and captivating area of research.

More specifically, in the incubation period, the articles were published in the HRM or other field such as accounting and finance, where, over the next decade, the management and business domain became the goal of 20 percent of the articles published. In the period of substantial growth, studies begin to be published also in the engineering field (4%) and 10 percent of the articles belong to other areas. In total, 40 percent of the research was published in HR management journal, while 35 percent was published in management and business journal. The evidence suggests that, while in the past the human capital was considered as a field of interest only for the human resources department, in recent years it has become increasingly more important and also managers and the business have understood its key role for the organization's success. Furthermore, the interest of the field of HR analytics has widened also in the field of engineering or in the information science area (included in "other"), which are constantly developing new techniques and methods to analyse HR data.

In the first period, the studies are for the most part conceptual. Then, during the phase of incremental growth, the number of empirical articles increased from 1 percent to 13 percent.

Finally, during the substantial growth time span, the empirical articles are the prevailing (25%), followed by the conceptual articles (24%). The growth of the empirical approach reflects the increasing interest in the use of data-driven approach that characterized this phase, and this trend is in line with the expansion of the HR analytics research in the engineering and information science fields.

The results in the geographical regions of the studies indicate that there is a transition over time upon which HR analytics research focuses. This geographical shift was also reported by Chalutz (2019). While in the incubation period the few articles on the topic of HR analytics and workforce measurement that specified a geographical region were published in North America and Oceania, in the following periods the research focuses also on the European and on the Asiatic countries.

TABLE 1
HR analytics and workforce measurement research characteristics by period of publication

Year of publication	1996-2000 (4 years) Incubation period n=7 (8%)	2008-2016 (8 years) Incremental growth n=30 (33%)	2017-2020 (3 years) substantial growth n=54 (59%)	1996-2020 total (24 years) n=91 (100%)
Type of journal				
HRM	5 (5%)	8 (9%)	27 (30%)	40 (44%)
Management and business	0	18 (20%)	14 (15%)	32 (35%)
Engineering	0	0	4 (4%)	4 (4%)
Other	2 (2%)	4 (4%)	9 (10%)	15 (16%)
Research cluster				
Empirical	1 (1%)	13 (14%)	23 (25%)	37 (41%)
Conceptual	5 (5%)	14 (15%)	22 (24%)	41 (45%)
Case based	1 (1%)	2 (2%)	6 (7%)	9 (10%)
Technical	0	1 (1%)	3 (3%)	4 (4%)
Geographical region (when indicated)				
North America	1 (1%)	0	4 (4%)	5 (5%)
Europe	0	2 (2%)	5 (5%)	7 (8%)
Asia	0	4 (4%)	4 (4%)	8 (9%)
Africa	0	2 (2%)	0	2 (2%)
Oceania	1 (1%)	3 (3%)	0	4 (4%)

Notes: Values=Number of articles; value in brackets=% of articles.
The articles of 2004 have been included in the incubation period.

The following sections analyze the results of the review.

Results

This chapter discusses and describes the most important themes that emerged from our analysis of the papers in the dataset.
To achieve a potential understanding of the scope and the areas of application of HR analytics in the organizations, we have arranged the 91 articles into 20 themes, as can be seen from the right part of the Table 3. Next, we investigated the evolution of the themes in the three different periods previously identified (Table 2).

The state of change of the HR analytics field pointed out in the reporting section of this paper is also proved by the shift of the focus in the themes in the various phases, as shown in the Table 2. In the first period, the studies have begun to examine prevalently how the use of metrics and the workforce measurement generate value to the organization (3%), and influence organizational productivity, performance and success (2%). Then, during the phase of incremental growth, the literature focused mainly on how the use of metrics and analytics help the HR function to demonstrate its value-add (4%) and on the contribution of the measurement of the HR operate on its professionalization (4%). The increase of the articles that examine the migration of the HR function toward a more strategic function reflects the awareness of the importance of the function also outside the HR domain, with the consequent increasing in the growth of the interest in publishing in the management and business field. Another field in which the research has focused is on how HR analytics and metrics help in strategic workforce management and planning. This topic covered 6 percent of the articles, which 4 percent were published during this period. Finally, during the substantial growth time span, the literature addressed predominantly the application of HR analytics to the management practices (8%) and the opportunities and challenges of HR analytics (8%).

TABLE 2
Topics characteristics by period of publication

Year of pubblication	1996-2000 (4 years) Incubation period n=7 (8%)	2008-2016 (8 years) Incremental growth n=30 (33%)	2017-2020 (3 years) substantial growth n=54 (59%)	1996-2020 total (24 years) n=91 (100%)
Themes				
ROI-based value creation;	0	2 (2%)	1 (1%)	3 (3%)
Value proposition;	0	0	1 (1%)	1 (1%)
Value generation;	3 (3%)	1 (1%)	3 (3%)	7 (8%)
Organizational productivity, performance and success;	2 (2%)	2 (2%)	5 (5%)	9 (10%)
Competitive advantage;	0	2 (2%)	1 (1%)	3 (3%)
Demonstrating HR value-add;	0	4 (4%)	3 (3%)	7 (8%)
Measurement contribution to the professionalism;	1 (1%)	4 (4%)	0	5 (5%)
Competence required;	0	0	3 (3%)	3 (3%)
Support in decision making;	1 (1%)	2 (2%)	6 (7%)	9 (10%)
Strategic workforce management and planning;	0	4 (4%)	2 (2%)	6 (7%)
Management practices;	0	1 (1%)	7 (8%)	8 (9%)
Adoption issues;	0	1 (1%)	4 (4%)	5 (5%)
Scholar-practitioner collaboration;	0	0	1 (1%)	1 (1%)
Opportunities/facilitators and challenges/barriers;	0	0	7 (8%)	7 (8%)
Aspects and uses;	0	2 (2%)	1 (1%)	3 (3%)

Employees' experience and well-being;	0	1 (1%)	2 (2%)	3 (3%)
Employee attitudes and behaviours;				
Ethics and trust;	0	0	1 (1%)	3 (3%)
Techniques and methods;	0	1 (1%)	1 (1%)	2 (2%)
	0	2 (2%)	3 (3%)	5 (5%)
Tools and technologies.	0	0	1 (1%)	1 (1%)
Notes: Values=Number of articles; value in brackets=% of articles. The articles of 2004 have been included in the incubation period.				

We grouped the themes into 6 topics: i) HR analytics and business value; (ii) the role and the evolution of the HR function (in creating value for the organization); (iii) applications of HR analytics, (iv) adoption and implementation of HR analytics practices; (v) organizational climate, culture and values; (vi) foundations of a data-driven HRM (Table 3).

TABLE 3

HR Analytics topics, themes and sample articles

Topics	**Themes and Sample Articles**
HR analytics and business value	**ROI-based value creation:** Chalutz, 2019; Steen & Welch, 2011; McNulty & Cieri, 2013. **Value proposition:** Tursunbayeva *et al*., 2018. **Value generation:** Sen & Haque, 2016 (internal stakeholders); Murphy & Zandvakili, 2000 (internal and external stakeholers); Boudreau & Ramstad, 1997; Vidgen *et al*., 2017; Boudreau, 1998; Van der Togt & Rasmussen, 2017; Kassic, 2019. **Organizational productivity, performance and success:** Mclver *et al*., 2018; Chhinzer & Ghatehorde, 2009; Du Plessis & De Wet Fourie, 2016; Poisat & Mey, 2017; Yeung & Berman, 1997; Ulrich, 1997; Mishra *et al*., 2018; Hazarika *et al*., 2019; Schiemann *et al*., 2018. **Competitive advantage:** Minbaeva, 2018; Nienaber & Sewdass, 2016; Manuja & Ghosh, 2014.
The role and the evolution of the HR function (in creating value for the organization)	**Demonstrating HR value-add:** Angrave *et al*., 2016; Kryscynski *et al*., 2017; Magau & Roodt, 2010; Pilenzo, 2009; Van den Heuvel & Bondarouk, 2017; Claus, 2019; Patre, 2016. **Measurement contribution to the professionalism:** Srimannarayana, 2010; Amalou-Döpke & Süß, 2014; Toulson & Dewe, 2004; Ulrich & Dulebohn, 2015; Tootell *et al*., 2009. **Competence required:** De Mauro *et al*., 2018; Sripathi 2018; Martin-Rios *et al*., 2017.
Applications of HR analytics	**Support in decision making:** Meyers *et al.*, 2019; Wingard, 2019; Dulebohn & Johnson, 2013; Levasseur, 2015; Boudreau, 1996; Sousa *et al*., 2019; Jabir *et al*., 2019; Noack, 2019, Kakkar, 2019. **Strategic workforce management and planning:** Huselid, 2018; Iwu *et al*., 2016; Wang & Cotton, 2017; Mulla & Premarajan, 2008, Akthar, 2013; Peiseniece & Volkova, 2010. **Management practices:** Wickramasinghe & Fonseka, 2012; Berhil *et al*., 2020; Garcia-Arroyo & Osca, 2019; Bekken, 2019; Sharma & Sharma, 2017; Nicolaescu *et al*., 2019; Herington *et al*.,2013; Kraichy & Schmidt, 2020; Frederiksen, 2017.

Adoption and implementation of HR Analytics practices	**Adoption issues:** Marler & Boudreau, 2017; Lismont *et al*., 2017; Aral *et al*., 2012; Dahlbom *et al*., 2019; Singh & Malhotra; 2020. **Scholar-practitioner collaboration:** Simon & Ferreiro, 2018. **Opportunities/facilitators and challenges/barriers:** Hamilton *et al*., 2019; Vargas *et al*., 2018; Bhardwaj & Patnaik, 2019; Andersen, 2017; Boudreau & Cascio, 2017; Levenson & Fink, 2017; Levenson, 2018. **Aspects and uses:** Gandhi & Minhaj, 2017; Narula, 2015; Handa & Garima 2014.
Organizational climate, culture and values	**Employees' experience and well-being:** Boyd & Gessner, 2013; Chattopadhyay *et al*., 2017; Greasley & Thomas, 2020. **Employee attitudes and behaviours:** Shah *et al*., 2017. **Ethics and trust:** Khan & Tang, 2016; Calvard & Jeske, 2018.
Foundations of a data-driven HRM	**Techniques and methods:** Gelbard *et al*., 2018; Safarishahrbijari, 2018; Wang & Katsamakas, 2019; Pape, 2016; Sinha *et al*., 2012 **Tools and technologies:** Florkowski, 2018.

The following part of this paper moves on to describe in greater detail what emerged through the examination of the literature.

Description of the HR analytics topics

HR analytics and business value

This stream of research primarily focuses on the way in which HR analytics plays a crucial role in firms' ability to achieve a competitive advantage (Chalutz, 2019).
The body of research in this topic is the first in terms of size (23 articles, over 25% of the articles analyzed).
A first group of articles delineates the ROI-based view of HR analytics and aims to provide scientific evidences to justify the efforts required to adopt analytic methods (Chalutz, 2019). They argue that ROI is an important measurement tool that may assist stakeholders in managerial decision making (Chalutz, 2019). They have shown that there is a connection between the HR investment in analytics and organization effectiveness, measured through an increased level of ROI (Chalutz, 2019).
A second group of articles focuses on the HR analytics value proposition and value generation. Their major goal is to provide insights into the value addition that HR metrics and analytics produce for the company. They address the ways in which the financial dimension of value associated with intangible assets creates sustainable long term advantages (Sen and Haque, 2016; Boudreau and Ramstad, 1997). The strategic role of high performance HR systems emerged as having an economically meaningful effect on firm level measures of financial performance and business outcomes (Sen and Haque, 2016; Tursunbayeva *et al*., 2018; Boudreau and Ramstad, 1997). Data and metric driven HRM help the unit to directly link its interventions to customers and employees, connecting them to the strategic value addition and making it visible to the organization (Murphy and Zandvakili, 2000, Sen and Haque, 2016; Boudreau and Ramstad, 1997; Boudreau, 1998; Van der Togt and Rasmussen, 2017; Mclver *et al*., 2018). This happens because the creation of good measures enables to learn how people influence productivity, customer and financial performance (Schiemann *et al*., 2018).

A third group of articles focuses on how the adoption of HR metrics and analytics affects the organizational productivity and financial performance. The relationship between HR metrics and organizational financial performance is supported by several studies (Chhinzer and Ghatehorde, 2009; Yeung and Berman, 1997; Hazarika *et al*., 2019), and it has been suggested that more translation of this relationship into HRM practices or policies is necessary (Chhinzer and Ghatehorde, 2009). It has been therefore demonstrated that the quality of HR practices is positively related to firm results (Ulrich, 1997). Many recent studies in this category (Poisat & Mey, 2017) have shown that there is an increased productivity in organizations that adopt electronic human resource management (e-HRM). This is seen as being achieved through the reduction of costs and administrative burden due to automation, that allows time for higher value tasks (Poisat & Mey, 2017). Moreover, e-HRM can help to improve the internal customers-service orientation of HR professionals (Poisat & Mey, 2017).
The last group of articles of this theme delineates the ways in which HR analytics support organizations in achieving competitive advantage. They argue that workforce metrics and the intelligence gained from analytics play an important role in the effective utilization, development, allocation and alignment of workers and their competence, and in developing competence, though rational decisions to ensure competitive advantage (Nienaber & Sewdass, 2016). Moreover, they enable to extrapolate data-driven informations, which leads to knowledge that help to make wise decisions that became part of the strategic plans that allow the organization to be one step ahead of its competitors (Du Plessis & De Wet Fourie, 2016; Chattopadhyay *et al*., 2017).

The role and the evolution of the HR function (in creating value for the organization)

This topic includes 15 (almost 17%) of the articles analyzed and focuses on the way in which HR analytics practices can be a powerful way for HR functions to add value to their organizations.
The amount of the research in this domain reflects the growing awareness that people are the most valuable asset for an organization's investment (Srimannarayana, 2010; Boudreau, 1998; Du Plessis and De Wet Fourie, 2016; Nienaber and Sewdass, 2016; Chattopadhyay *et al*., 2017) and the greatest source of competitive advantage (Tootell *et al*., 2009; De Mauro *et al*., 2018; Boudreau, 1998), and the effective human capital management is critical to an organisation's success (Magau and Roodt, 2010; Pilenzo, 2009; Srimannarayana, 2010; Amalou-Döpke and Süß, 2014, Toulson and Dewe, 2004). However, HR professionals had always difficulties in demonstrating the value of their function and in being recognized as a key contributor (Kryscynski *et al*., 2017; Pilenzo, 2009). One explanation is that HR fall short of being a strategic partner because it lacks evidence-based rigor in decision making (Kryscynski *et al*., 2017). Several studies suggest that improving HR measurement and data-driven HR systems is a powerful way to help the HR function to explain the value of its profession and move toward a more business and strategic orientation (Amalou-Döpke and Süß, 2014). Furthermore, a relationship between high commitment management practices and the financial performance of organizations has been established, supporting the view that measuring the accomplishment of the HR function is important because managing human resources does lead to tangible returns (Toulson & Dewe, 2004).
A number of authors (Lawler, Levenson and Boudreau, 2004, cited in Magau and Roodt, 2010; Sripathi, 2018; Amalou-Döpke & Süß, 2014; Martin-Rios *et al*., 2017; Sen and Haque, 2016; Kassic, 2019; Mclver *et al*., 2018; Chhinzer and Ghatehorde, 2009; Du Plessis and De Wet Fourie, 2016; Yeung & Berman, 1997; Hazarika *et al*., 2019; Chattopadhyay *et al*., 2017) showed that the use of metrics and analytics increased the scope of HR, which can demonstrate its value-add though measuring its human capital contribution, particularly in terms that are understood by stakeholders (Tootell *et al*., 2009), migrating it from a more administrative to a more strategic discipline. Similarly,

Patre (2016) states that HR analytics is a powerful tool for organizations in optimizing their workforce decisions for best business results. In the same vein, Heuvel and Boundarouk (2016) argues that sophisticated HR practices are not only a powerful way for organization as a whole to add value increasing performance and attitudes, but also for HR to add value to their function that become more strategic and find out what it means to be "strategic". Overall, as Claus (2019) points out, there seems to be some evidence to indicate that the new talent management value proposition is evolving from talent acquisition to designer or architect of worker experiences and the HR function is shifting from being an administrative maintenance function to being view as a core business function that could contribute to organizational effectiveness (Ulrich and Dulebohn, 2015). Improved HR analytics are a means to helping codify and make value happen (Ulrich and Dulebohn, 2015).
Despite these evidences, the HR department hasn't yet adopted the analytics solutions as much as other business functions such as R&D, accounting, operations, marketing or finance and so forth (Heuvel and Boundarouk, 2016; Magau and Roodt, 2010; Kryscynski *et al*., 2017; Tootell *et al*., 2009). Therefore, it is necessary to educate the HR in scientific thinking and statistical reasoning, in order to advance the quality of people decision (Van der Togt and Rasmussen, 2017).

Applications of HR analytics

The applications of HR analytics is a central topic and is the first in term of the size together with the "HR analytics and business value". It is discussed by the 26 percent of the articles in the dataset.

A first group of articles addresses the way in which data and models assist employees and managers to make decisions and it observes that the number of managers and employees that adopts metrics and analytics to help solve key HR problems is increasing (Dulebohn and Johnson, 2013).
These articles and their recent peak in growth are the result of an increasing call for adopting technology to support decision-making at a variety of organizational levels and for a variety of purposes (Dulebohn and Johnson, 2013), thus giving growing attention to increasing the quality of data, improving the integration of data from different sources, and providing the means to examine a variety of alternative scenarios (Dulebohn and Johnson, 2013).
Boudreau (1996) suggests that measures allows decision makers to make more rational and productive choices. Similarly, Sousa *et al.* (2019), Jabir (2019), Noack (2019), Wingard (2019) and Kakkar (2019) claim that analytics lead organization in making better decision making and with more accuracy.
In conclusion, these studies show that providing accurate, timely and relevant data to support HR decision making is one of the central purposes of the organizations nowadays (Dulebohn and Johnson, 2013).
A second group of articles focuses on how HR analytics improve strategic workforce management and planning. These studies pay particular attention to how HR analytics assists leaders and managers in managing the workforce more effectively, enabling them to achieve their operational and strategic objectives (Huselid, 2018).
They suggest that the use of HR analytics provide important insight on how differentiate and manage a differentiated workforce (Wang and Cotton, 2017).
One example of the above is represented by the study of Wang and Cotton (2017, cited in Huselid 2018) in which workforce analytics has been applied to investigate the differences in the contribution of strategic and support roles. They highlighted the importance of differentiating human resource management practices for the two different roles (Wang and Cotton, 2017).
A final group of articles describe the role of HR analytics in human resource management practices.

The strategic role of big data in this area is particularly visible in the processes of recruitment and selection, organizational development and knowledge management, analyzing behaviuors, fostering security, health and well-being, and increasing efficiency (Garcia-Arroyo & Osca, 2019).
In particular, the role of HR analytics on performance management is investigated in three of the articles included in the dataset (Sharma and Sharma, 2017; Nicolaescu et al., 2019; Herington *et al*., 2013).
Sharma and Sharma (2017) recognize that the use of HR analytics in the performance appraisal system help to take fact-based insights, reducing the subjectivity bias and enhancing employees' perceived accuracy and fairness. They argue that this perception of accuracy and fairness further increases employees' satisfaction with the performance appraisal system, and thus employees' willingness to improve performance. Nicolaescu *et al.* (2019) propose a scoring algorithm that calculates employee's value within the organization. The study of Herington *et al.* (2013) highlights the challenge of quantifying HR achievements to determine performance effects.
Two studies attempted to examine the turnover using organizational data (Kraichy and Schmidt, 2020; Frederiksen, 2017). Using various HR metrics, Kraichy and Schmidt (2020) have been able to investigate the collective turnover at different job levels. In the same vein, Frederiksen (2017) predicted employee quits analyzing personnel and job satisfaction surveys data.

Adoption and implementation of HR Analytics practices

Studies in this subject describe the process of implementation of HR analytics practices and the challenges that organizations face.
The body of research in this topic is the second in terms of size (16 articles).
The first subtopic examines the adoption issues with HR analytics. They build on the assumption that HR analytics is at the early adopters stage (Marler and Boudreau, 2017; Lismont *et al.*, 2017; Andersen, 2017).
Despite several lines of evidence suggest that being able to derive insights from data has become more and more important in the last few years (Lismont *et al.*, 2017), as described on the previous topics, there are various reasons for why HR analytics is not yet widespread within organizations. Lismont *et al.* (2017) identifies the strictness of privacy regulations and the hardness to collect the right data as the major causes of the low adoption rate. Furthermore, HR professionals have been not familiar with working with a data-driven approach (Dahlbom *et al.*, 2019). Dahlbom *et al.* (2019) identifies the absence of skills necessary for the HR function for operating in data-driven way, together with the lack of business understanding, outdated IT infrastructure and systems, difficulties in moving beyond reporting and misconceptions regarding big data and its utility for HR as the most common obstacles emerged in the adoption of HR analytics. Similarly, Marler and Boudreau (2017) lists three significant moderators that impact the relationship between the adoption of HR analytics and organizational impact. These are: providing employees with the right knowledge and skills, building a network of supportive stakeholders and ensuring the quality and accessibility of the data and capabilities of e-HRM software system.
The second subtopic includes only the study of Simon and Ferreiro (2018), who used a case study approach to investigate the collaboration between practitioners and researchers in the process of development of a workforce analytics initiative. As this case very clearly demonstrates, the cooperation between the two parts is vital for developing innovation and creating alternative ways of approaching problems, analyzing data, and presenting the work to other departments (Simon and Ferreiro, 2018).
In the third subtopic, researchers have explored the major challenges and opportunities that organizations face in the implementation of HR analytics. Their major goal is to identify factors that

facilitates, or act as barriers to the adoption of HR analytics. In a Delphi method study conducted by Bhardwaj and Patnaik (2019), respondents reported that the number of challenges seems to be more than opportunities, as the organizations lack skilled manpower. This study confirms the evidence from the previous observations (Simon and Ferreiro, 2018) that suggest that a scholar-practitioner collaboration could be helpful to fill each other's gaps in both sides. Bhardwaj and Patnaik (2019) identified internal politics, resistance to change, alignment of overall objectives of the company and collaboration of traditional method as the challenges in the area.
Vargas *et al.* (2018) identified attitude toward analytics and technology and quantitative self-efficacy, followed by social influence and trialability as the main driver of the analytics adoption process.
Finally, the last subtopic assesses the aspects and uses of HR analytics and includes only three articles. The main utilizations of HR analytics that emerged are: develop workforce skills in key areas, identify and develop high potential employees, understand and plan for future talent needs, design and build career paths for valued employees, create a pipeline of successors for high performers, plan and measure outcomes of executive leadership development (Narula, 2015), measuring the employee sentiment, improving employee satisfaction and discovery and underlying reasons for attrition (Minhaj and Gandhi, 2017).

Organizational climate, culture and values

This topic defines the role that HR analytics perform in making positive changes for the workforce.
Only 6 articles belong to this subject and they are mainly empirical. Despite the practical relevance of this subject, the small number of papers may be an indication that this issue is a research frontier and the literature is still in its infancy. It includes the contribution of HR analytics to address issues such as high attrition, employment branding, work-life balance, congenial reporting relationships (Chattopadhyay *et al*., 2017); to reevaluate the performance metrics in order to consider the well-being of employees (Boyd & Gessner, 2013); to be aware of well-being (Greasley & Thomas, 2020) and to assess employee attitudes and behaviors (Shah *et al*., 2017).
This area of research is close to the corporate social responsibility debate on the role of HR in the implementation of socially responsible interventions and in promoting equity and social justice for employees (Boyd & Gessner, 2013), as well as on the role of HR practices in supporting employee motivation and engagement as part of organizational change and readiness programmes (Shah *et al*., 2017). In their empirical study, Shah *et al*. (2017) describe how, by analyzing data of talents, managers are able to understand their attitudes and behaviors, and through the assessment of individuals they can drive the organizational change programmes.
The contributions of Chattopadhyay *et al*. (2017) and Greasley and Thomas (2020) delineates the effects of data, measures and analytics from the employees' perspective. They address the way in which such practices support the employees' experience and well-being and they argue that in order to make HR analytics more effective, a more "user focused" perspective is required.
As noted by Khan and Tang (2016) and by Calvard & Jeske (2018), while there is a growing adoption of sophisticated techniques for analyzing employee data for helping organizations to enhance both operational efficiencies and competitive postures, the issue of the perception of the HR analytics from the employee perspective and the impact of such perception on employee outcomes has received little attention. It is also important to consider this aspect because "the employee attribution of managerial motives behind HR practices impact important employee-, group-, and organizational-level outcomes" (Khan and Tang, 2016).
Deepen this research area could help organizations to enhance a positive employee attribution of HR analytics practices (Khan and Tang, 2016) and to develop appropriate HR practices for dealing

with big data challenges and preventing risks to individuals, employers, the organization at large as well as external stakeholders like data storage providers and consultants (Calvard and Jeske, 2018). This positive perception benefits the employees, the organization and the stakeholders oriented toward them (Khan and Tang, 2016).

Foundations of a data-driven HRM

Studies in this area focuses on the data analytics techniques applied to HR analytics. Similarly to the previous topic, the small amount of papers in this domain (only 6) may be an indication that this issue is a research frontier. Papers in this topic foreshadow a lot of potential roles of algorithms, methods and open source tools that allow the HR function to gain data-driven rich insights. These recent contributions recognize that the increasing availability of detailed data and the development of sophisticated data science algorithm enable organizations to understand how they really work, and they benefit from technology for organizing and managing work (Wang and Katsamakas, 2019).
The first subtheme includes almost all articles and explores a number of different, advanced data analytics techniques and methods that can be used to analyze people data. Papers in this subtheme presents five different approach to evaluate human factors: a conceptual ontology (Gelbard et al., 2018), the workforce modeling and prediction methods (Safarishahrbijari, 2018), a network data science approach (Wang & Katsamakas, 2019), a prescriptive framework to prioritise data items for business analytics (Pape, 2016) and the social media analytics practices (Sinha et al., 2012).
These studies highlight several benefits in analyzing data with these methods instead of collecting them through traditional surveys. Firstly, the assessment of employee perception through traditional tools such as periodical survey-based data collection is expensive and often inflated and biased (Gelbard *et al*., 2018). Data mining procedures, instead, allow organizations to evaluate human factors in a real-time, convenient and more reliable way (Gelbard *et al*., 2018). Secondly, the use of detailed data and the development of sophisticated data science algorithm enable the organizations to understand how they really work and how technology affects performance (Wang and Katsamakas, 2019). Finally, the vast range of operation research technique ensures that each one adapts to every single different specific need (Safarishahrbijari, 2018).
The second subtheme contains only the article of Florkowski (2018). In his chapter, he defined the content domains and provide an assessment of the HR technologies, with the aim to understand their impacts.

Discussion and Implications

The aim of the present research was to review the major topics and themes of the HR analytics and workforce measurement research, focusing on its purpose, applications and value production to identify where the field has progressed and where it has not, to provide recommendations for moving this research forward. We conduct a systematics literature review from 1996 to 2020. This study has provided a deeper insight into the main characteristics of the literature and the major topics and theme addressed. Two dimensions/constructs emerged as key feature of the topics addressed in the HR analytics and workforce measurement literature: the prevalent HR orientation dimension ("people" *versus* "processes" *versus* "technology") and the prevalent focus of HR analytics/measurement initiatives ("internal" *versus* "external"). The two dimensions provide the basis for a typology of HR analytics and workforce measurement topics classification through which HR measurement and analytics might contribute to create value for the organization and the society as a whole (see Figure 3). The proposed typology contributed to the understanding of the areas of focus in the literature. These results are similar to those reported by De Stefano *et al.* (2017).

Compared to their classification, we have found another dimension: the “technology” one. The results of the review highlights that the topics and the dimensions are related to each other: the technology influences the processes, which in turn influence people. The internal and external effects of the HR analytics and measurement practices appeared to mutually influence each other. Future research should be undertaken to investigate these interrelationships more in detail.
The articles classified in the “external processes” and “internal technology” quadrant represent the majority of the sample and reflect the dominance of the instrumental role of HR in creating value and in being applied within the organisations. The internal focus, that represents 61 percent of the literature, is slightly in majority compared to the external. The articles classified in the “external technology” and “external people” quadrant represent the minority of the sample. Although HR analytics and measurement practices play a primary role in managing overall employees’ experience, well-being, attitudes, behaviours and trust, very few articles addressed this issue. To capture this subject, future research can examine how HR analytics practices influence the organizational climate and culture and the ethical implications of HR analytics.
Although the HR analytics data analysis techniques, methods, tool and technologies are the foundation for analyze people data for improve the organizational processes and the success, we found a small number of papers on this topic. Most publications on these topics were found in other disciplinary areas and were rejected in the executing phase because did not refer to applications in the HR area. To move forward, future research can build on the literature on data analytics tools and techniques to understand how they can be applied in the HR area.

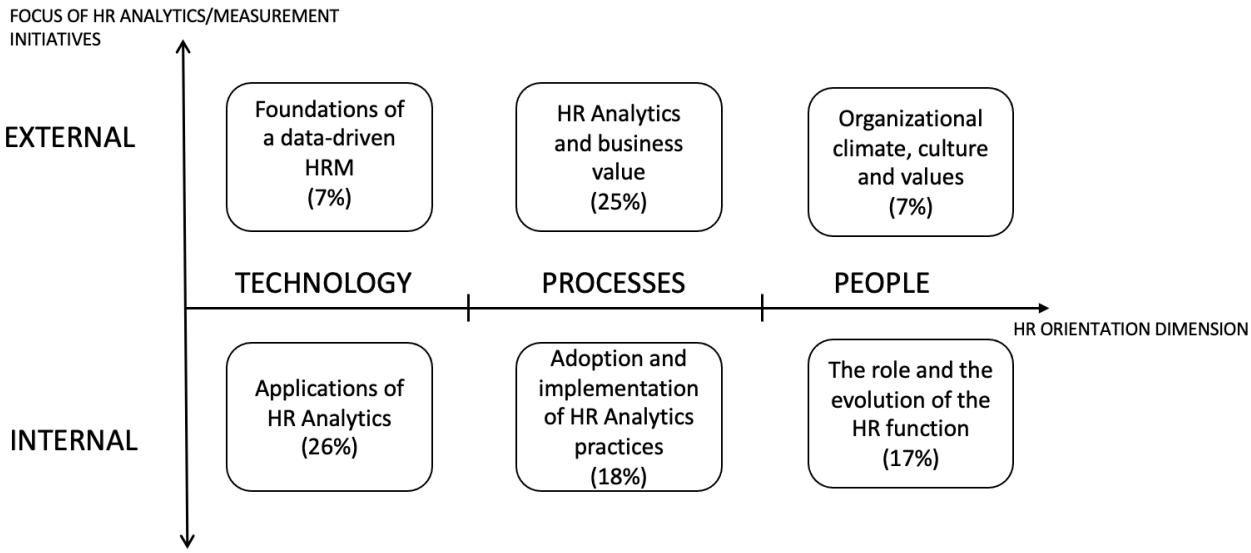


Figure 3: HR analytics and workforce measurement topics classification

Conclusions

We reviewed the articles published in scholarly peer-reviewed journals until the first months of 2020 to identify how HR analytics and workforce measurement research evolved, the major topics and themes and progresses over time, to find areas where where progress is lacking and to investigate what are the main purposes and applications of HR metrics and analytics and how these produce value for the organization. We proposed a framework for classifying and understanding the potential role HR analytics and measurement play in the various organizational and external contexts. We used the findings to identify directions for future research aimed toward future understanding of the interrelationships between the topics and dimensions more in detail and toward investigating areas which we found to be still under-explored.

The study is limited by the lack of information on how the HR analytics research is affected by the current health emergency caused by the spread of the Covid-19, because the collection of paper stopped before the beginning of the pandemic. This would be a fruitful area for further work.

References

Akthar, P. (2013). "Importance of measuring HR's effectiveness: a drive to HR metrics". *International Journal of Research in Commerce & Management*. *4*(10):78-81.

Amalou-Döpke, L., & Süß, S. (2014). HR measurement as an instrument of the HR department in its exchange relationship with top management: A qualitative study based on resource dependence theory. *Scandinavian Journal of Management*, *30*(4), 444–460. https://doi.org/10.1016/j.scaman.2014.09.003

Andersen, M. K. (2017). Human capital analytics: The winding road. *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 133–136. https://doi.org/10.1108/JOEPP-03-2017-0024

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology. *Management Science*, *58*(5), 913–931. https://doi.org/10.1287/mnsc.1110.1460

Bekken, G. (2019). The Algorithmic Governance of Data driven-Processing Employment: Evidence-based Management Practices, Artificial Intelligence Recruiting Software, and Automated Hiring Decisions Social Sciences, Sociology, Management and complex organizations. *Psychosociological Issues in Human Resource Management*, *7*(2), 25. https://doi.org/10.22381/PIHRM7220194

Berhil, S., Benlahmar, H., & Labani, N. (2020). A review paper on artificial intelligence at the service of human resources management. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*(1), 32. https://doi.org/10.11591/ijeecs.v18.i1.pp32-40

Bhardwaj, S., & Patnaik, S. (2019). People Analytics: Challenges and Opportunities - A Study Using Delphi Method. *IUP Journal of Management Research, 18*(1), 7–23.

Boudreau, J. (1996). The Motivational Impact of Utility Analysis and HR Measurement. *Journal of Human Resource Costing & Accounting*, *1*(2), 73–84. https://doi.org/10.1108/eb029031

Boudreau, J. (1998). Strategic Human Resource Management Measures: Key Linkages and the PeopleVantage Model. *Journal of Human Resource Costing & Accounting,* Vol. 3 Iss 2 pp. 21 – 40. http://dx.doi.org/10.1108/eb029046

Boudreau, J., & Cascio, W. (2017). Human capital analytics: Why are we not there? *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 119–126. https://doi.org/10.1108/JOEPP-03-2017-0021

Boudreau, J., & Ramstad, P. M. (1997). Measuring Intellectual capital: Learning from Financial History. *Cornell - Center for Advanced Human Resource Studies, Papers*, *36*.

Boyd, N., & Gessner, B. (2013). Human resource performance metrics: Methods and processes that demonstrate you care. *Cross Cultural Management: An International Journal*, *20*(2), 251–273. https://doi.org/10.1108/13527601311313508

Calvard, T. S., & Jeske, D. (2018). Developing human resource data risk management in the age of big data. *International Journal of Information Management*, *43*, 159–164. https://doi.org/10.1016/j.ijinfomgt.2018.07.011

Chalutz Ben-Gal, H. (2019). An ROI-based review of HR analytics: Practical implementation tools. *Personnel Review*, *48*(6), 1429–1448. https://doi.org/10.1108/PR-11-2017-0362

Chattopadhyay, D., Biswas, D. D., & Mukherjee, S. (2017). A New Look at HR Analytics. *Globsyn Business School*, 11, 41-51.

Chhinzer, N., & Ghatehorde, G. (2009). Challenging Relationships: HR Metrics and Organizational Financial Performance. *The Journal Of Business Inquiry, 8*(1), 37-48.

Claus, L. (2019). HR disruption—Time already to reinvent talent management. *BRQ Business Research Quarterly*, *22*(3), 207–215. https://doi.org/10.1016/j.brq.2019.04.002

Dahlbom, P., Siikanen, N., Sajasalo, P., & Jarvenpää, M. (2019). Big data and HR analytics in the digital era. *Baltic Journal of Management*, *15*(1), 120–138. https://doi.org/10.1108/BJM-11-2018-0393

De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, *54*(5), 807–817. https://doi.org/10.1016/j.ipm.2017.05.004

De Stefano, F., Bagdadli, S., & Camuffo, A. (2017). The HR role in corporate social responsibility and sustainability: A boundary-shifting literature review. *Human Resource Management*, *57(2), 549-566*. https://doi.org/10.1002/hrm.21870

Du Plessis, A. J., & De Wet Fourie, L. (2016). Big Data and HRIS used by HR practitioners: Empirical evidence from a longitudinal study. *Journal of Global Business & Technology*, *12*(2), 44–55.

Dulebohn, J. H., & Johnson, R. D. (2013). Human resource metrics and decision support: A classification framework. *Human Resource Management Review*, *23*(1), 71–83. https://doi.org/10.1016/j.hrmr.2012.06.005

Florkowski, G. W. (2018). HR Technology Systems: An Evidence-Based Approach to Construct Measurement. In M. R. Buckley, A. R. Wheeler, & J. R. B. Halbesleben (Eds.), *Research in Personnel and Human Resources Management* (Vol. 36, pp. 197–239). Emerald Publishing Limited. https://doi.org/10.1108/S0742-730120180000036006

Frederiksen, A. (2017). Job satisfaction and employee turnover: A firm-level perspective. *German Journal of Human Resource Management: Zeitschrift Für Personalforschung*, *31*(2), 132–161. https://doi.org/10.1177/2397002216683885

Garcia-Arroyo, J., & Osca, A. (2019). Big data contributions to human resource management: A systematic review. *The International Journal of Human Resource Management*, 1–26. https://doi.org/10.1080/09585192.2019.1674357

Gelbard, R., Ramon-Gonen, R., Carmeli, A., Bittmann, R. M., & Talyansky, R. (2018). Sentiment analysis in organizational work: Towards an ontology of people analytics. *Expert Systems*, *35*(5), e12289. https://doi.org/10.1111/exsy.12289

Greasley, K., & Thomas, P. (2020). HR analytics: The onto-epistemology and politics of metricised HRM. *Human Resource Management Journal*, 1748-8583.12283. https://doi.org/10.1111/1748-8583.12283

Hamilton, R. H., & Sodeman, W. A. (2020). The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources. *Business Horizons*, *63*(1), 85–95. https://doi.org/10.1016/j.bushor.2019.10.001

Handa, D., & Garima (2014). Human Resource (HR) Analytics: emerging trend in HRM (HRM). *CLEAR International Journal of Research in Commerce & Management*, *5*(6), 59–62.

Hazarika, I., Albeshr, M., & Cho, B. (2019). Role of HR metrics in enhancing firm performance of selected UAE airline companies. *Academy of Strategic Management Journal, 18*(6), 8.

Herington, C., McPhail, R., & Guilding, C. (2013). The evolving nature of hotel HR performance measurement systems and challenges arising: An exploratory study. *Journal of Hospitality and Tourism Management*, *20*, 68–75. https://doi.org/10.1016/j.jhtm.2013.06.002

Huselid, M. A. (2018). The science and practice of workforce analytics: Introduction to the *HRM* special issue. *Human Resource Management*, *57*(3), 679–684. https://doi.org/10.1002/hrm.21916

Iwu, C., Kapondoro, L., Twum-Darko, M., & Lose, T. (2016). Strategic Human Resource Metrics: A Perspective of the General Systems Theory. *Acta Universitatis Danubius : Oeconomica*, *12*, 5–24.

Jabir, B., Falih, N., & Rahmani, K. (2019). HR analytics a roadmap for decision making: Case study. *Indonesian Journal of Electrical Engineering and Computer Science*, *15*(2), 979. https://doi.org/10.11591/ijeecs.v15.i2.pp979-990

Kakkar, H., & Kaushik, S. (2019). Technology Driven Human Resource Measurement–A Strategic Perspective. *International Journal on Emerging Technologies, 10*(1a): 179-184.

Kassick, D. (2019). "Workforce Analytics and Human Resource Metrics: Algorithmically Managed Workers, Tracking and Surveillance Technologies, and Wearable Biological Measuring Devices," *Psychosociological Issues in Human Resource Management* 7(2): 55–60. doi:10.22381/PIHRM7120199

Khan, S. A., & Tang, J. (2017). The paradox of human resource analytics: Being mindful of employees. *Journal of General Management, 42*(2):57-66. doi:10.1177/030630701704200205

Kraichy, D., & Schmidt, J. (2019). Collective turnover: Organization design and processes or contagion effects? *Employee Relations: The International Journal*, *42*(2), 492–506. https://doi.org/10.1108/ER-01-2019-0055

Kryscynski, D., Reeves, C., Stice-Lusvardi, R., Ulrich, M., & Russell, G. (2017). Analytical abilities and the performance of HR professionals: Analytical Ability in HR Professionals. *Human Resource Management*, *57*(3), 715–738. https://doi.org/10.1002/hrm.21854

Levasseur, R. E. (2015). People Skills: Building Analytics Decision Models That Managers Use—A Change Management Perspective. *Interfaces*, *45*(4), 363–364. https://doi.org/10.1287/inte.2015.0798

Levenson, A. (2018). Using workforce analytics to improve strategy execution. *Human Resource Management*, *57*(3), 685–700. https://doi.org/10.1002/hrm.21850

Levenson, A., & Fink, A. (2017). Human capital analytics: Too much data and analysis, not enough models and business insights. *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 145–156. https://doi.org/10.1108/JOEPP-03-2017-0029

Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, *37*(3), 114–124. https://doi.org/10.1016/j.ijinfomgt.2016.12.003

Magau, M. D., & Roodt, G. (2010). An evaluation of the Human Capital BRidge™ framework. *SA Journal of Human Resource Management*, *8*(1). https://doi.org/10.4102/sajhrm.v8i1.276

Manuja, S., & Ghosh, O. (2014). A Study on Importance of Strategic Human Resource Management. *Amity Global HRM Review*, *4*, 9–17.

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, *28*(1), 3–26. https://doi.org/10.1080/09585192.2016.1244699

Martin-Rios, C., Pougnet, S., & Nogareda, A. M. (2017). Teaching HRM in contemporary hospitality management: A case study drawing on HR analytics and big data analysis. *Journal of Teaching in Travel & Tourism*, *17*(1), 34–54. https://doi.org/10.1080/15313220.2016.1276874

McIver, D., Lengnick-Hall, M. L., & Lengnick-Hall, C. A. (2018). A strategic approach to workforce analytics: Integrating science and agility. *Business Horizons*, *61*(3), 397–407. https://doi.org/10.1016/j.bushor.2018.01.005

McNulty, Y., & Cieri, H. D. (2013). Measuring Expatriate Return on Investment with an Evaluation Framework. *Global Business and Organizational Excellence*, *32*(6), 18–26. https://doi.org/10.1002/joe.21511

Meyers, T. D., Vagner, L., Janoskova, K., Grecu, I., and Grecu, G. (2019). Big Data-driven Algorithmic Decision-Making in Selecting and Managing Employees: Advanced Predictive Analytics, Workforce Metrics, and Digital Innovations for Enhancing Organizational Human Capital Social Sciences, Sociology, Management and complex organi. *Psychosociological Issues in Human Resource Management*, *7*(2), 49. https://doi.org/10.22381/PIHRM7220198

Minbaeva, D. B. (2018). Building credible human capital analytics for organizational competitive advantage. *Human Resource Management*, *57*(3), 701–713. https://doi.org/10.1002/hrm.21848

Minhaj, M., & Gandhi, L. (2017). A Big Deal for C-Suite in Talent Management. FOCUS International Journal of Management, ISSN: 0973 9165 12 (2).

Mishra, D., Luo, Z., Hazen, B., Hassini, E., & Foropon, C. (2018). Organizational capabilities that enable big data and predictive analytics diffusion and organizational performance: A resource-based perspective. *Management Decision*, *57*(8), 1734–1755. https://doi.org/10.1108/MD-03-2018-0324

Mulla, Z. R., & Premarajan, R. K. (2008). Strategic Human Resource Management in Indian it Companies: Development and Validation of a Scale. *Vision: The Journal of Business Perspective*, *12*(2), 35–46. https://doi.org/10.1177/097226290801200204

Murphy, T.E., & Zandvakili, S. (2000). Data- and metrics-driven approach to human resource practices: Using customers, employees, and financial metrics. *Human Resource Management*, *39*(1), 93–105. Scopus. https://doi.org/10.1002/(SICI)1099-050X(200021)39:1<93::AID-HRM8>3.0.CO;2-N

Narula, S. (2015). HR Analytics: its use, techniques and impact. *CLEAR International Journal of Research in Commerce & Management*, *6*(8), 47–52.

Nicolaescu, S. S., Florea, A., Kifor, C. V., Fiore, U., Cocan, N., Receu, I., & Zanetti, P. (2020). Human capital evaluation in knowledge-based organizations based on big data analytics. *Future Generation Computer Systems*, *111*, 654–667. https://doi.org/10.1016/j.future.2019.09.048

Nienaber, H., & Sewdass, N. (2016). A reflection and integration of workforce conceptualisations and measurements for competitive advantage. *Journal of Intelligence Studies in Business*, *6*(1). https://doi.org/10.37380/jisib.v6i1.150

Noack, B. (2019). “Big Data Analytics in Human Resource Management: Automated Decision-Making Processes, Predictive Hiring Algorithms, and Cutting-Edge Workplace Surveillance Technologies Social Sciences, Sociology, Management and complex organizations. *Psychosociological Issues in Human Resource Management*, *7*(2), 37. https://doi.org/10.22381/PIHRM7220196

Pape, T. (2016). Prioritising data items for business analytics: Framework and application to human resources. *European Journal of Operational Research*, *252*(2), 687–698. https://doi.org/10.1016/j.ejor.2016.01.052

Patre, S. (2016). Six Thinking Hats Approach to HR Analytics. *South Asian Journal of Human Resources Management*, *3*(2), 191–199. https://doi.org/10.1177/2322093716678316

Pilenzo, R. (2009). A New Paradigm for HR. *Organization Development Journal*, 27(3), 63-75.

PLoS Medicine (OPEN ACCESS) Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Poisat, P., & Mey, M. R. (2017). Electronic human resource management: Enhancing or entrancing? *SA Journal of Human Resource Management*, *1*(2). https://doi.org/10.4102/sajhrm.v15i0.858

Safarishahrbijari, A. (2018). Workforce forecasting models: A systematic review. *Journal of Forecasting*, *37*(7), 739–753. https://doi.org/10.1002/for.2541

Schiemann, W. A., Seibert, J. H., & Blankenship, M. H. (2018). Putting human capital analytics to work: Predicting and driving business success: Putting human capital analytics to work. *Human Resource Management*, *57*(3), 795–807. https://doi.org/10.1002/hrm.21843

Sen, A., & Haque, S. (2016). HR Metrics and the Financial Performance of a Firm. *Journal of Management Research (09725814)*, *16*(3), 177–184.

Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, *70*, 366–378. https://doi.org/10.1016/j.jbusres.2016.08.010

Sharma Anshu, & Sharma Tanuja. (2017). HR analytics and performance appraisal system: A conceptual framework for employee performance improvement. *Management Research Review*, *40*(6), 684–697. https://doi.org/10.1108/MRR-04-2016-0084

Simón, C., & Ferreiro, E. (2018). Workforce analytics: A case study of scholar-practitioner collaboration: Workforce Analytics and scholar-practitioner collaboration. *Human Resource Management*, *57*(3), 781–793. https://doi.org/10.1002/hrm.21853

Singh, T., & Malhotra, S. (2020). Workforce Analytics: Increasing Managerial Efficiency in Human Resource. *International Journal of Scientific & Technology Research, 9*(1), 3260-3266.

Sinha, V., Subramanian, K. S., Bhattacharya, S., Chaudhuri, K. (2012). “The contemporary framework on social media analytics as an emerging tool for behavior informatics, HR analytics and business process”. Journal of Contemporary Management Issues. Vol. 17, 2012, 2, pp. 65-84.

Sousa, M.J., Pesqueira, A. M., Lemos, C., Sousa, M., & Rocha, Á. (2019). Decision-Making based on Big Data Analytics for People Management in Healthcare Organizations. *Journal of Medical Systems*, *43*(9). Scopus. https://doi.org/10.1007/s10916-019-1419-x

Srimannarayana, M. (2010). Status of HR Measurement in India. *Vision: The Journal of Business Perspective*, *14*(4), 295–307. https://doi.org/10.1177/097226291001400406

Sripathi, K., & Madhavaiah, D. C. (2018). Are HR Professionals Ready to Adopt HR Analytics? A Study on Analytical Skills of HR Professionals. *Control Systems*, *10*, 6.

Steen, A., & Welch, D. (2011). Are Accounting Metrics Applicable to Human Resources? The Case of Return on Investment in Valuing International Assignments. *Australasian Accounting, Business and Finance Journal,* 5(3), 57-72.

Stuart, M., Angrave, D., Charlwood, A., kirkpatrick, ian, & Lawrence, M. (2016). HR and Analytics: Why HR is set to fail the big data challenge. *Human Resource Management Journal*, *26(1):1-11*. https://doi.org/10.1111/1748-8583.12090/

Tootell, B., Blackler, M., Toulson, P., & Dewe, P. (2009). Metrics: HRM’s Holy Grail? A New Zealand case study. *Human Resource Management Journal*, *19*(4), 375–392. https://doi.org/10.1111/j.1748-8583.2009.00108.x

Toulson, P. K., & Dewe, P. (2004). HR accounting as a measurement tool. *Human Resource Management Journal*, *14*(2), 75–90. https://doi.org/10.1111/j.1748-8583.2004.tb00120.x

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*, 207–222. https://doi.org/10.1111/1467-8551.00375.

Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018). People analytics—A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management*, *43*, 224–247. https://doi.org/10.1016/j.ijinfomgt.2018.08.002

Ulrich, D. (1997). Measuring human resources: An overview of practice and a prescription for results. *Human Resource Management*, *36*(3), 303–320. https://doi.org/10.1002/(SICI)1099-050X(199723)36:3<303::AID-HRM3>3.0.CO;2-#

Ulrich, D., & Dulebohn, J. H. (2015). Are we there yet? What's next for HR? *Human Resource Management Review*, *25*(2), 188–204. https://doi.org/10.1016/j.hrmr.2015.01.004

van den Heuvel, S., & Bondarouk, T. (2017). The rise (and fall?) of HR analytics: A study into the future application, value, structure, and system support. *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 157–178. https://doi.org/10.1108/JOEPP-03-2017-0022

van der Togt, J., & Rasmussen, T. H. (2017). Toward evidence-based HR. *Journal of Organizational Effectiveness: People and Performance*, *4*(2), 127–132. https://doi.org/10.1108/JOEPP-02-2017-0013

Vargas, R., Yurova, Y. V., Ruppel, C. P., Tworoger, L. C., & Greenwood, R. (2018). Individual adoption of HR analytics: A fine grained view of the early stages leading to adoption. *The International Journal of Human Resource Management*, *29*(22), 3046–3067. https://doi.org/10.1080/09585192.2018.1446181

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626–639. https://doi.org/10.1016/j.ejor.2017.02.023

Wang, L., & Cotton, R. (2018). Beyond *Moneyball* to social capital inside and out: The value of differentiated workforce experience ties to performance: Beyond moneyball to social capital inside and out. *Human Resource Management*, *57*(3), 761–780. https://doi.org/10.1002/hrm.21856

Wang, N., & Katsamakas, E. (2019). A Network Data Science Approach to People Analytics: *Information Resources Management Journal*, *32*(2), 28–51. https://doi.org/10.4018/IRMJ.2019040102

Wickramasinghe, V., & Fonseka, N. (2012). Human resource measurement and reporting in manufacturing and service sectors in Sri Lanka. *Journal of Human Resource Costing & Accounting*, *16*(3), 235–252. https://doi.org/10.1108/14013381211286388

Wingard, Devin (2019). "Data-driven Automated Decision-Making in Assessing Employee Performance and Productivity: Designing and Implementing Workforce Metrics and Analytics," *Psychosociological Issues in Human Resource Management* 7(2): 13–18. doi:10.22381/PIHRM7120192.

Yeung, A. K., & Berman, B. (1997). *Adding value through human resources: Reorienting human resource measurement to drive business performance. Human Resource Management, 36(3), 321–335.* doi:10.1002/(sici)1099-050x(199723)36:3<321::aid-hrm4>3.0.co;2-y.

People Analytics: a Case Study on Predicting Employee Attrition Using Machine Learning Techniques

Abstract

Recently, there has been renewed interest in adopting data analytics to help solve HR problems and to make more informed and effective choices. One of the greatest challenges for organizations is employee turnover because of its adverse impact in many areas, such as organization's image, productivity and performance. To address this issue, machine learning tools have been developed for investigating and predicting employee attrition, as well as methods for evaluating their predictive power. Evidence on what are the most important predictors that lead to attrition and in what areas it is more likely to happen enable HR managers to implement targeted retention policies and practices. However, dealing with the complexity of data from HR Information Systems (HRIS) and choosing the best classifiers suitable for the case of interest, which makes each project particular, is a critical task. This study is about predicting employee attrition using machine learning models on a real dataset of a large Italian financial company. This contrasts with much extant research which is based on artificial datasets. The contribution of this paper is to explore and compare the performance of several common models which are found in the extant literature on real data. We then focus on the results of the best performing model, and identify some groups of employees who have a high risk of attrition on which the company could arrange interventions to reduce voluntary resignation.

Introduction

In recent years, there is a growing awareness that human resources (HR) are one of the most critical assets for the success of an organization and for a company to be competitive. For this reason, various studies have been conducted to improve the process of recruitment and selection, to help organizations make better personnel selection decisions (Chien and Chen, 2008). The employees represent a real investment for organizations (Fallucchi et al., 2020) and when an employee leaves the company, the organization may be losing a valuable resource, thus facing (temporary) impairment of processes, replacement and recruiting costs (Fallucchi et al., 2020; Negassa, 2016; AlSayyed and Braiki, 2015; Prabakaran and Vetrivel, 2017; Zylka, 2016; Zylka and Fischbach, 2017; Varadharaj and Irfan, 2019; Getachew, 2017; Workagegn, 2017; Balcha, 2019; Staw, 1980; Abeble, 2016; Imani, 2013). Employees' attrition has other negative consequences, such as demoralization of extant employees (Zylka, 2016; Negassa S. N., 2016; Zylka and Fischbach, 2017; Varadharaj and Irfan, 2019; Staw, 1980; Hussein, 1989; Ayuure, 2013; Abeble, 2016; Imani, 2013) increased work-load (Abeble, 2016), negative reputational effects for the organization, unfulfilled daily functions (Varadharaj and Irfan, 2019), costs related to the effort of finding replacements, both direct and indirect, i.e. opportunity costs, (Prabakaran and Vetrivel, 2017); encouraging other employees to look for better job opportunities (Zylka, 2016; Zylka and Fischbach, 2017), loss of employees' confidence in the company (Getachew, 2017), further unexpected leaves and remaining co-workers having to compensate by, as well as re-allocation of resources (Zylka and Fischbach, 2017, Varadharaj and Irfan, 2019), customer service failures (Varadharaj and Irfan, 2019; Abeble, 2016); management frustration (Varadharaj and Irfan, 2019; Abeble, 2016), loss of experiences and knowledge (Getachew, 2017; Hussein, 1989, Abeble, 2016), loss of output because of reduced

workforce and training costs of newly hired people (Balcha, 2019). Therefore, reducing employees' turnover benefits companies in several ways and is an important challenge (Chang, 2009).

In the literature, several terms are used to describe employees' resignations, the most common of which are "attrition" and "turnover". Attrition is defined by Nappinnai and Premavathy (2013: 11) as ‹‹a gradual reduction in workforce without firing of personnel, as and when workers resign or retire and are not replaced››. For Negi (2013), attrition means the inevitable loss of workforce for any reason. Singh and Singh (2017: 3) define turnover as ‹‹the act of leaving a current job such as moving to another job or relocating to another destination››. Similarly, the term "turnover" is used by AlSayyed and Braiki (2015: 649) to refer to ‹‹the movement of employees in and out of employment with respect to a given organization or company››. In the same vein, Zylka and Fischbach (2017: 54) uses the term "turnover behavior" to refer to ‹‹leaving a job for a similar, alternative job across organizational boundaries››. A further definition of turnover is given by Arokiasamy (2013: 1532) who describes it as ‹‹the entrance of new employees into the organization and the departure of existing employees from the organization››. Then, "attrition" and "turnover" are similar terms, and both occur when an employee leaves the company; "attrition" tends to be used to refer to the loss of employees as a naturally occurring event (such as retirement, attending school, raising a family) while "turnover" applies to employees who leave the company due to issues related with it (McQuerrey, 2019). In this article, we use the term "attrition" to mean "voluntary resignation".

Understanding which employees are likely to leave and why is crucial for organizations, since the specific knowledge of the reasons that lead to attrition helps in developing policies and strategies for employee retention (Alao and Adeyemo, 2013). Recently, the growing digitalization of relationship models between company and employees makes an increasing number of information available to HR functions, which can be enhanced through new tools and methodologies to improve the attractiveness, evaluation and development of human capital. Despite the efforts of HR managers to adopt metrics and analytics to help solve key HR problems is also increasing, there has been little quantitative analyses that investigate and predict employees' attrition. Lismont and co-authors (2017) identify strict privacy regulations and difficulties in collecting right data as the major causes of the low adoption rate. Furthermore, HR professionals have not been familiar with working with a data-driven approach (Dahlbom et al., 2019). Simon and Ferreiro (2018) used a case study approach to investigate the collaboration between practitioners and researchers in developing workforce analytics initiatives; they conclude that their cooperation is vital for developing innovation and creating alternative ways of approaching problems, analyzing data, and presenting the work to other company departments than HR.

In this article, we present the outcome of our investigation of a real data set from a large Italian company aimed at predicting attrition and identifying in what areas it is mostly significant and what employees' features are associated with it. By "large" we mean that the company has several branches, business lines and many employees (more than 5,000); also, its branches are distributed across several Italian geographical regions. Our analyses are useful for the HR department to take timely and appropriate employee retention policies.

The article is structured as follows. First, a literature review is carried out to better identify predictors of employee turnover. Second, the case study is discussed, describing the methodology and data analysis. Third, several machine learning techniques are described and implemented. Finally, we assess the performance of each model and discuss the results of the best one.

1. Theoretical Framework

Literature review on the analysis of the causes of attrition and turnover

Introduction and definition of terms

The purpose of the literature review is to explore existing research into the causes of attrition. Several definitions of attrition and turnover have been proposed. Employee attrition is often confused with employee turnover.

Attrition is defined by Nappinnai and Premavathy (2013: 11) as "a gradual reduction in workforce without firing of personnel, as and when workers resign or retire and are not replaced". For Negi (2013), attrition means the inevitable loss of workforce for any reason.
Singh and Singh (2017: 3) define turnover as "the act of leaving a current job such as moving to another job or relocating to another destination". Similarly, the term 'turnover' is used by AlSayyed and Braiki (2015: 649) to refer to "the movement of employees in and out of employment with respect to a given organization or company". In the same vein, Zylka and Fischbach (2017: 54) uses the term 'turnover behavior' to refer to "leaving a job for a similar, alternative job across organizational boundaries". A further definition of turnover is given by Arokiasamy (2013: 1532) who describes it as "the entrance of new employees into the organization and the departure of existing employees from the organization".

The two terms are similar, and both occur when an employee leaves the company. In the literature, the term 'attrition' tends to be used to refer to the loss of employees as a naturally occurring event (such as retirement, attending school, raising a family…). Employee turnover, on the other hand, applies to employees who leave the company due to issues related with it (McQuerrey, 2019).

Having defined what is meant by attrition and turnover, we will now move on to discuss the literature review methodology and results.

Method

Search and collect the documents:

Database: Scopus

Query definition:

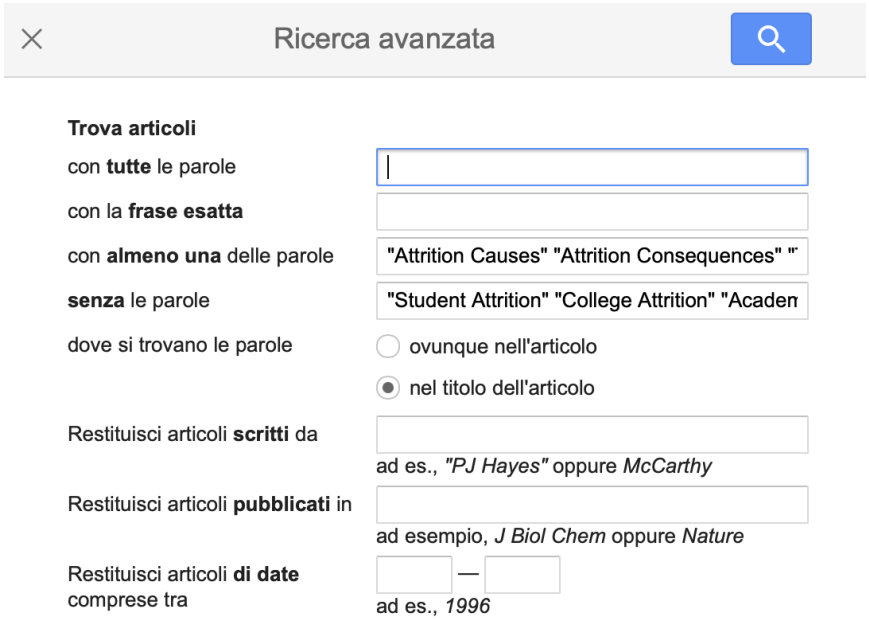
TITLE-ABS-KEY (***"Attrition Causes"*** OR ***"Attrition Consequences"*** OR ***"Turnover Causes"*** OR ***"Turnover Consequences"***) AND TITLE-ABS (***"Employee"*** OR ***"Organization"***) AND NOT (***"Student Attrition"*** OR ***"College Attrition"*** OR ***"Academic Attrition"*** OR ***"Student Turnover"*** OR ***"College Turnover"*** OR ***"Academic Turnover"***)

Results: 9
Record excluded because of access restrictions: 3
Final records: 6

Database: Google Scholar

Query definition:
con almeno una delle parole (***"Attrition Causes" "Attrition Consequences" "Turnover Causes" "Turnover Consequences"***)
senza le parole (***"Student Attrition" "College Attrition" "Academic Attrition" "Student Turnover" "College Turnover" "Academic Turnover"***)

(allintitle: "Attrition Causes" OR "Attrition Consequences" OR "Turnover Causes" OR "Turnover Consequences" -"Student Attrition" -"College Attrition" -"Academic Attrition" -"Student Turnover" -"College Turnover" -"Academic Turnover")



Results: 80
Record excluded because of access restrictions: 10
Record excluded because not in English (foreign language): 2
Final records: 68

Other records identified through an initial search: 12

Screening of the papers: PRISMA Flow Chart

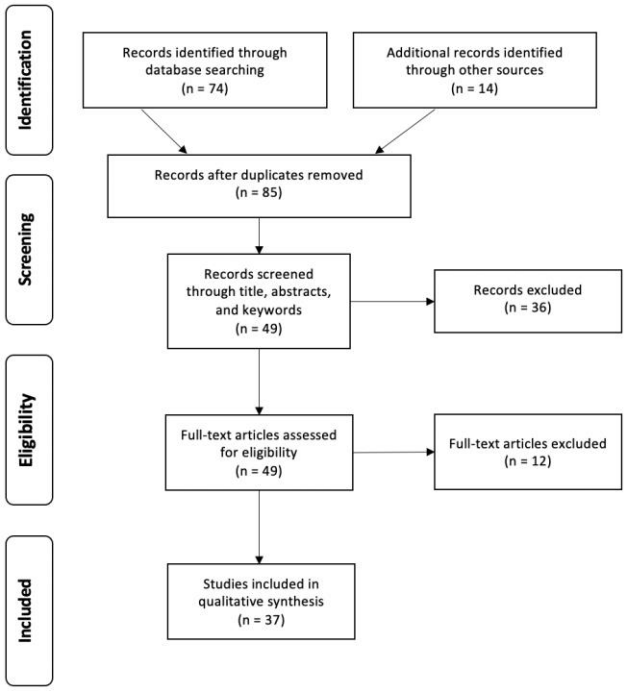


Fig. 1: Prisma flow chart visualising the article selection process

Results of the literature review

Internal causes:

Compensation and benefits practices: insufficient salary/satisfaction with pay (Colding, 2015; Negi, 2013; Getachew, 2016; Wang and Chen, 2013; Arokiasamy, 2013; Varadharaj and Irfan, 2019; Getachew, 2017; Workagegn, 2017; Hussein, 1989; Ayuure, 2013; Abeble, 2016; Zhang *et al.*, 2018; Selhadin, 2019; Tuji, 2013), delay in payment, no/delayed increment, wage compression (Negi, 2013), benefits and indirect compensation (pension plans, vacation pay based on length of service) (Bennett *et al.*, 1993; Getachew, 2016; Getachew, 2017; Ayuure, 2013), fringe benefits (Arokiasamy, 2013; Varadharaj and Irfan, 2019; Selhadin, 2019), contributing significantly to the organization but having wages fall short of the current market rate (Prabakaran and Vetrivel, 2017).

Organizational tenure (Cohen, 1993; Balcha, 2019; Hussein, 1989).

Tenure in role (Taylor *et al.*, 1996).

Relationships: cooperation, treatment, fairness (perceived organizational justice), tolerance, helpfulness, the style of assigning and performing tasks (Rhodes, 1983), lack of teamwork (Parker and Skitmore, 2005), romantic relationships at work (Zhang A., 2019), morale/motivational problems with project team and staff (Parker and Skitmore, 2005), supervisor satisfaction, poor supervision (Woo and Maertz, 2012; Wang and Chen, 2013; Hussein, 1989; Selhadin, 2019) co-worker satisfaction (Woo and Maertz, 2012), lack of respect (Imani, 2013), mobbing/harassment, supervisors or co-workers leaving the organization (Woo and Maertz, 2012), attrition of the group members (Woo and Maertz, 2012), poor relationship between employees and management (Prabakaran and Vetrivel, 2017; Varadharaj and Irfan, 2019), dissatisfaction among new workers caused by inadequate selection and assignment method (Hussein, 1989).

Performance and potential (Wigert, 2018; Woo and Maertz, 2012).

Recognition: prestige, opportunities, development (Negassa, 2016), promotional opportunities (Colding, 2015), feeling unappreciated (Parker and Skitmore, 2005; Workagegn, 2017; Wang and Chen, 2013), underutilization of talents and skills of the individuals (Nappinnai & Premavathy, 2013), lack of recognition (Rhodes, 1983; Nappinnai and Premavathy, 2013; Abeble, 2016), low social status (Wang and Chen, 2013), lack of internal incentive mechanism (Zhang *et al.*, 2018), lack of responsibility (Imani, 2013).

Expectations: role ambiguity (Colding, 2015; Rhodes, 1983), imbalance between work and personal life, unclear assignments (role ambiguity), expectations, perceived organizational support (POS), inadequate information on how to execute the job adequately, blurred expectations of peers and supervisors, vague performance evaluation techniques, extensive job pressures, absence of agreement on job functions or duties (Rhodes, 1983).

Promotion and growth opportunities: biased promotion, delayed promotion (Negi, 2013; Arokiasamy, 2013; Varadharaj and Irfan, 2019), less or no career growth opportunities/lack of advancement opportunities/lack of career progression (Nappinnai and Premavathy, 2013; Parker and Skitmore, 2005; Negi, 2013; AlSayyed and Braiki, 2015; Prabakaran and Vetrivel, 2017; Wang and Chen, 2013; Arokiasamy, 2013; Workagegn, 2017; Hussein, 1989; Ayuure, 2013; Abeble, 2016; Zhang *et al.*, 2018; Selhadin, 2019; Tuji, 2013), professional stagnation/lack of career development (Parker and Skitmore, 2005; Getachew, 2016; Getachew, 2017).

Communication: type of communication, feedback, sincerity, awareness, information asymmetry and respecting opinions (Negassa, 2016), ethics (Rhodes, 1983; Parker and Skitmore, 2005), lack of communication (AlSayyed and Braiki, 2015).

Organizational culture: workload, flexible working hours, access to sources, type of culture, poor personnel policies, poor recruitment policies, poor grievance procedures, or lack of motivation (Rhodes, 1983), fit with company culture (Parker and Skitmore, 2005), poor supervisory practices (Rhodes, 1983; Nappinnai and Premavathy, 2013; Wang and Chen, 2013; Hussein, 1989), lack of freedom of expression (Nappinnai and Premavathy, 2013), discordance between enterprise culture and personal lifestyle (Zhang *et al.*, 2018).

Organizational changes: dissatisfaction of organizational changes and restructuring (AlSayyed and Braiki, 2015).

Transfer: forceful transfer, transfer to a place employee is not willing to go (Negi, 2013).

Physical/technical workplace environment: lack of hygiene, lack of basic facilities like water, canteen (Negi, 2013), dangerous work conditions (Woo and Maertz, 2012; Akinyomi, 2016; Getachew, 2016), workstation set-up, furnitures/work equipment design and quality, tools, facilities, restrooms, suitable lighting, proper ventilation, health and safety provisions (Akinyomi, 2016; Prabakaran and Vetrivel, 2017; Getachew, 2016; Getachew, 2017; Selhadin, 2019), proper equipment, machinery and computer technology, work space and ergonomically-correct seating (Workagegn, 2017), design and age, workplace layout, temperature, space, noise, radiation, air quality (Getachew, 2017).

Psychological workplace environment: stressful environment (Akinyomi, 2016; Prabakaran and Vetrivel, 2017); isolating working environment (that lowers hinders employees from interacting freely) (Getachew, 2016), privacy (Singh and Singh, 2017).

Organizational factors: size (number of employee), location, nature kind, stability (Balcha, 2019), bad company policies (Ayuure, 2013), inadequate of corporate regulations (Zhang *et al.*, 2018), organizational commitment (Imani, 2013).

Training and orientation: lack of proper training and supervision/poor training handling system/inadequate training and development program (that do not goes to specific needs of employees) (AlSayyed and Braiki, 2015; Getachew, 2016; Workagegn, 2017), lack of well-organized training programs (Hussein, 1989), poor orientation (Imani, 2013).

Tasks: monotony of tasks, task-labour mismatch, team issues, lack of job autonomy (Negi, 2013; Abeble, 2016), task repetitiveness/boredom (Colding, 2015; Abeble, 2016).

Leadership style: leading to confusion related to directions and commands which generate frustration among the workforce (Negi, 2013), absence of empowerment (Abeble, 2016), discordance between career planning and enterprise development status (Zhang *et al.*, 2018)

Lack of flexibility: lack of flexibility in timing, lack of a flexible work schedule, choice of task, introduction of new technology and employees incompetency / unwillingness to learn and understand (Negi, 2013; Hochwarter et al., 1999); night shifts (Wang and Chen, 2013), lack of work-life balance (Abeble, 2016)

Lack of job security: fear of being expelled/ retrenched/ terminated. Faulty performance appraisal. Underestimation of performance. Power distance and politics. Communication gap between management and workforce (Negi, 2013).

External causes

Demographics characteristics: age, gender (e.g. females could have higher turnover rate due to the responsibilities of childcare) (Negassa, 2016; Bennett *et al.*, 1993; Wang and Chen, 2013; Balcha, 2019; Hussein, 1989; Abeble, 2016; Imani, 2013; Tuji, 2013), race (Negassa, 2016; Bennett *et al.*, 1993), experience (Abeble, 2016).

Personality attitudes and characteristics: personality traits (Woo and Maertz, 2012; Hussein, 1989), personal characteristics (Ayuure, 2013; Abeble, 2016), over-sensitivity (Negi, 2013), emotional state (Arokiasamy, 2013), unrealistic expectations (Balcha, 2019).

Private life events: marital status (Negassa, 2016; Hussein, 1989; Abeble, 2016; Tuji, 2013), partner separation or divorce (Woo and Maertz, 2012), end of life/ pregnancy, shift of family, death of family member (or in general mourning for a close person) (Woo and Maertz, 2012), retirement (El-Rayes *et al.*, 2020), children care/number of children (Wang and Chen, 2013; Balcha, 2019; Tuji, 2013), relocation of the partner/spouse (Abeble, 2016).

Physical and mental health conditions: medical conditions (El-Rayes *et al.*, 2020), mental imbalance (Negi, 2013); fatigue due to night shifts (Wang and Chen, 2013).

Educational level: educational level (workers with higher levels of educational achievement usually have higher expectations and competence levels) (Negi, 2013; Negassa, 2016; Bennett *et al.*, 1993; Hussein, 1989; Abeble, 2016; Varghese *et al.*, 2019), overeducation/overqualification (Woo and Maertz, 2012).

Individual preferences: location or community attachment (like or dislike their town, the location of the workplace, or the climate in their part of the country) (Woo and Maertz, 2012), wish to go abroad/to travel, wish to be self-employed (Negi, 2013; Abeble, 2016), wish to learn new skills (Workagegn, 2017), preferences, motives and needs (Arokiasamy, 2013).

Home-work distance (Sullivan, 2015).

External labour market: better pay (AlSayyed and Braiki, 2015; Negi, 2013; Ayuure, 2013), better working conditions (Ayuure, 2013), changes of promotion, better perks, more fringe and benefits in other organizations (Negi, 2013; AlSayyed and Braiki, 2015), eagerness to get into companies with global presence (Nappinnai and Premavathy, 2013), job opportunities/alternative job possibilities /perceived alternative job opportunities (Negassa, 2016; Parker and Skitmore, 2005; Wang and Chen, 2013; Arokiasamy, 2013; Varadharaj and Irfan, 2019).

Other social and economic factors: society's economic development level, labor market condition, employment system (Negassa, 2016; Bennett *et al.*, 1993; Workagegn, 2017; Ayuure, 2013), enterprise property, education and health care facilities around the organization, transportation services available around the organization, housing services, the cost of living, quality of life (Negassa, 2016), unionization (Arokiasamy, 2013; Varadharaj and Irfan, 2019; Hussein, 1989).

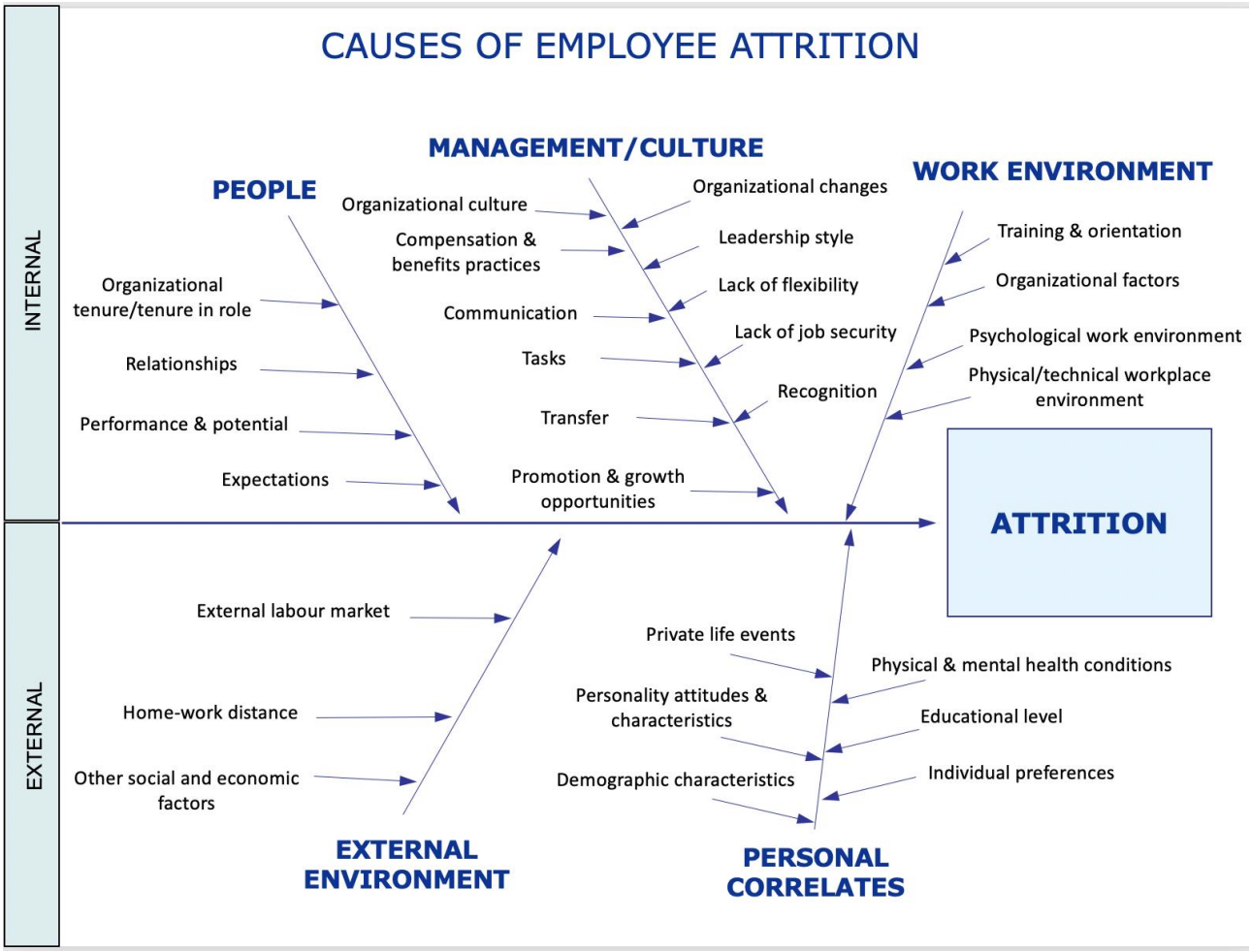


Fig. 2: Ishikawa diagram on the causes of attrition

Literature review on the analysis of the consequences of attrition

Negative consequences

Direct costs of turnover: replacement and recruiting costs (advertising, interviewing, testing…), training time and costs (Negassa, 2016; AlSayyed and Braiki, 2015; Prabakaran and Vetrivel, 2017; Zylka, 2016; Zylka and Fischbach, 2017; Varadharaj and Irfan, 2019; Getachew, 2017; Workagegn, 2017; Balcha, 2019; Staw, 1980; Abeble, 2016; Imani, 2013), leaving costs – payroll and HR administration (Balcha, 2019), selection and placement costs, orientation costs, lost wages/salaries, administrative costs, loss of human capital, and customer satisfaction problems (Negassa, 2016), time of the newer employee to be easy with the new system, with the co-employee, to be familiar with the new environment etc, vacancy problems (AlSayyed and Braiki, 2015), loss of productivity and performance (AlSayyed and Braiki, 2015; Negassa, 2016; Prabakaran and Vetrivel, 2017; Varadharaj and Irfan, 2019; Getachew, 2017), cost of inefficiency of the new staff (Prabakaran and Vetrivel, 2017).

Indirect costs of turnover: operational disruption, demoralization of employees (who remain) (Zylka, 2016; Negassa S. N., 2016; Zylka and Fischbach, 2017; Varadharaj and Irfan, 2019; Staw, 1980; Hussein, 1989; Ayuure, 2013; Abeble, 2016; Imani, 2013), increased work-load (Abeble, 2016), negative image on the organization, employee development plans fail (Negassa, 2016), unfulfilled

daily functions (Varadharaj and Irfan, 2019), hidden costs (instead of an organization expending substantial amount of money and time trying to find replacements for disengaged employees, it could have dedicated such resources and energy in productive activities that will contribute towards moving the organization in achieving its objectives) (Prabakaran and Vetrivel, 2017); relationships with co-workers negatively affected (Zylka, 2016; Zylka and Fischbach, 2017); interpretation of the departures of a former colleague as a rejection of the job (begin to realize the possibility that better job opportunities exist) (Zylka, 2016; Zylka and Fischbach, 2017), loss of confidence (in the project) (Getachew, 2017), unexpected leaves and remaining co-workers have to compensate by (Zylka and Fischbach, 2017), customer service (Varadharaj and Irfan, 2019; Abeble, 2016); management frustration (Varadharaj and Irfan, 2019; Abeble, 2016), distractions (Varadharaj and Irfan, 2019), loss of experiences and skilled personnel/loss of knowledge (Getachew, 2017; Hussein, 1989, Abeble, 2016), loss of output (from those leaving before they are replaced, because of delays in obtaining replacements, while new starters are on their learning curve) (Balcha, 2019).

Positive consequences

Enhance individual and organizational work performance: better job skills and more motivation/productivity of the new employee (Negassa, 2016; Zylka, 2016; Zylka and Fischbach, 2017; Staw, 1980)

Reduction of entrenched conflict: resolve conflicts (when a conflicting supervisor or co-worker leaves an organization) (Negassa, 2016; Zylka, 2016; Zylka and Fischbach, 2017; Staw, 1980)

Increasing mobility and morale: the turnover may open positions in an otherwise impenetrable hierarchy, being a creator of promotion opportunities (Negassa, 2016; Zylka, 2016; Zylka and Fischbach, 2017); "resolve deep-seated conflicts" between the conflicting parties and contributes to organizational morale (Zylka, 2016), termination of bad matches (Workagegn, 2017), creating opportunities for advancement (Hussein, 1989).

Positive attitudes: job satisfaction, organizational commitment (Zylka, 2016; Zylka and Fischbach, 2017).

Social capital gain: the newly arrived employee increases his or her social capital and experiences socialization through the new employment (Zylka, 2016; Zylka and Fischbach, 2017).

Setting the culture right (innovation and adaptation): import new type of knowledge, ideas, experience and skills (Negassa, 2016; Workagegn, 2017; Staw, 1980), introduction of change (Hussein, 1989).

Cost savings: leave of relatively expensive employees (Workagegn, 2017**;** Negassa, 2016), coping mechanism for individual under stress and invite absenteeism, carelessness, sabotage, and other non-productive behaviours (Hussein, 1989).

Data acquisition and understanding

The data include 27 features for each record of the employee and 5767 observations (we decided to omit "CodicePersona", the employee count variable, because it is a sequence of number and is not therefore important to understand attrition).
Table 1 shows the characteristics of the dataset.

Table 1: Human Resource Datasets Attributes

Concept	**Variable**	**Measurement Level**	**Missing values**
Dependent variable Attrition	Dimissioni	Nominal	0
Independent variables			
Gender	Genere	Nominal	0
Educational level	TitoloStudio	Ordinal	292
Working place of the organization	ClasseStruttura	Nominal	0
Detail of the cluster of the working place	DettaglioStruttura	Nominal	0
Role	Ruolo	Nominal	0
Position (staff/officer or responsible/manager)	Posizione	Nominal	0
Age at 31/12/2019	Eta	Ordinal	0
Organizational tenure	AnniServizio	Ordinal	0
Part time or full time	Orario	Nominal	0
Pay-grade	Grado	Ordinal	0
% hours worked (according to the contract)	QuotaOreLavoro	Ordinal	82
Type of national contract	Contratto	Nominal	0
Talent program career	Talento	Nominal	10
Bonus	Premio	Ordinal	0
Compensation	RalMensile	Ordinal	0
Retention compensation for apical roles	IndennizzoRetention	Ordinal	0
Occasional bonus	Gratifica	Ordinal	0
Home town	Citta	Nominal	276
Home address	Indirizzo	Nominal	276
Province of residence	Provincia	Nominal	276
Number of sons	NumeroFigli	Ordinal	82

Potential	Potenziale	Ordinal	933
Average market wage	RetribuzioneAbi	Ordinal	895
A score of leadership success used to evaluate high-ranks	Leadership	Ordinal	5139
Outcome of the (last) survey about employees satisfaction	Clima	Ordinal	177

Definition of the scope and out of scope: variable of the literature in the study of attrition and our variable comparison

Analysis of the causes

Based on the available data, we have identified a number of variables in our dataset that have some potential correspondence with one or more of the factors present in the literature. Fig. 3 and Table 1 summarize the correspondence between our variables and the factors that can be determinant in studying the phenomenon of attrition according to the literature.

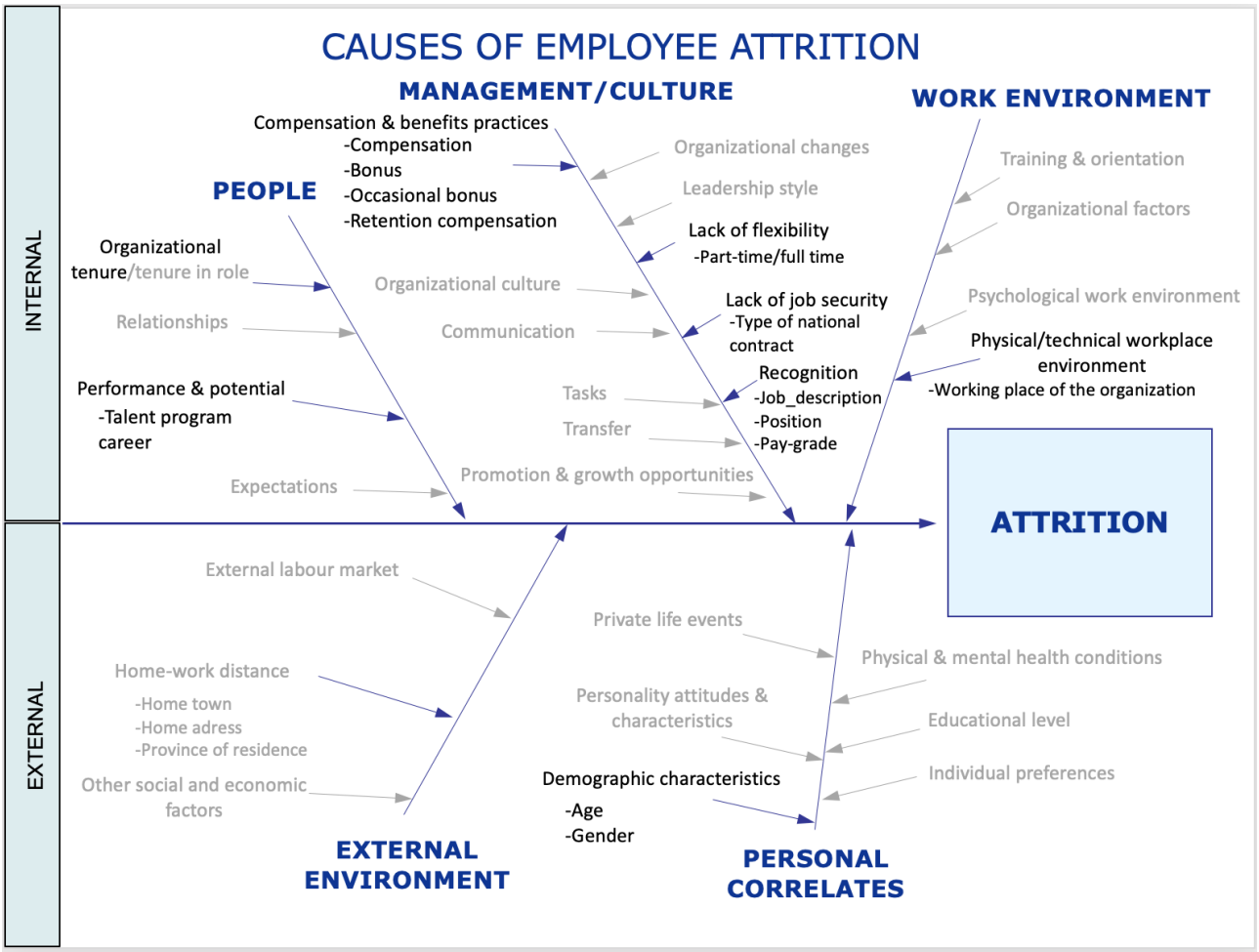


Fig. 3: Ishikawa diagram of the causes of attrition present in the dataset

Table 2: the positioning of the dataset's variables according to the literature

Scope	Factors	Dataset variables
People	Organizational tenure	Organizational tenure
	Performance & potential	Potential
		Talent Program career
Management/culture	Compensation & benefit practices	Compensation
		Bonus
		Occasional bonus
		Retention compensation for apical roles
	Communication	Leadership
		Clima
	Lack of flexibility	Part-time/full-time
	Lack of job security	Type of national contract
	Recognition	Role
		Position
		Pay-grade
Work environment	Organizational factors	Size
		Location
	Physical/technical workplace environment	Working place of the organization
		Detail of the cluster of the working place
External environment	External labour market	Average market wage
Personal correlates	Private life events	Number of sons
	Demographic characteristics	Age
		Gender
	Physical & mental conditions	% hours worked
	Educational level	Educational level

Description of the relevant characteristics for making predictions

Having analysed the causes of attrition and turnover, we will now move on to describe in greater detail the factors available in the data provided by the company. After describing each item, the following sections will investigate what are the characteristics and interactions between these characteristics most relevant to predict the phenomenon of attrition.

Demographic characteristics:

Age

A number of researchers have found an association between age-related differences and turnover rates, that are higher in work forces that are, on average, younger (Bennett *et al.*, 1993; Abeble, 2016; Bennett *et al.*, 1993; Hussein, 1989). One reason is that younger employees may be less likely to have personal constrains and mobility limitations and they perceive more alternative opportunities than older employees (Bennett *et al.*, 1993; Hussein, 1989). Moreover, it has been observed that younger workers have less family responsibilities, less financial and familiar compulsions that dictate to continue employment and so they are more likely to take risks (Hussein, 1989; Abeble, 2016; Imani, 2013). Finally, with increase in age a person has greater level of prestige, confidence and patience on the workplace (Imani, 2013; Tuji, 2013).

Gender

Research evidence suggests that females are more likely to leave the organization, and the gender-related differences can be explained by the fact that the women are in need to birth, take care of the family, have more responsibilities of childcare and thus more difficulties in balancing work and family life (Negassa, 2016; Wang and Chen, 2013; Tuji, 2013). Moreover, Bennett *et al.* (1993) argue that women are less likely to be employed by core firms where job offer security, relatively high returns on education and experience and are more often employed by periphery firms where wages and returns to human capital are lower.

Private life events:

Number of sons

Tuji (2013) suggests that employees who have children prefer to stay in organizational areas that they stabilized their family life. In addition, especially for women, childcare responsibilities are linked with the probability of leaving the organization (Wang and Chen, 2013).

Educational level:

Educational level has a significant impact on turnover rate, because higher levels of educational achievement contribute to the increase in the desire to find a working position that meets one's expectations and competences (Wang and Chen, 2013). Moreover, more educated employees are confident to have more external employment opportunities available to them (Abeble, 2016). In particular, young, inexpert (inexperienced) employees with a high level of education tend to have

low level of happiness about jobs and career, and lower commitment to the organization (Negassa, 2016).

Physical and mental health conditions:

% hours worked

Non-firm specific motivations, such as medical conditions, mental imbalance and fatigue due to night shifts can lead to attrition (El-Rayes *et al.*, 2020; Negi, 2013; Wang and Chen, 2013). Since absences from work being are often justified by physical or mental illness, we assume that the percentage of hours worked could be a proxy to measure this factor.

Home-work distance:

Sullivan (2015) argues that commute issues can have a negative impact on employee retention. According to Sullivan (2015) commute issues are associated with negative implications such as absenteeism, tardiness, turnover, decrease of productivity due to stress and reduced team and manager performance. What is expected therefore is that employees who need to travel more time and distance to reach their places of work will be more inclined to leave their job when a competitor offers them a similar job opportunity that is closer to where the worker live.

External labour market:

Average market wage

It has been observed that the availability of higher paying jobs is one of the most common reasons for leaving the current occupation (AlSayyed and Braiki, 2015; Negi, 2013; Ayuure, 2013).

Performance and potential:

Potential and talent program career

The more talented employees have the higher expectations of their workplaces and they are also more likely to have other opportunities available to them (Wigert, 2018).

Organizational tenure:

As noted by Cohen (1993), when tenure increase, the individual's investments increase, and the cost of leaving can lead to higher levels of organizational commitment. Cohen (1993) investigated the relationship between age and tenure and organizational commitment, but he has shown that this latter is in relation to important organizational outcomes such as turnover. Similarly, Hussein (1989) argues that tenure is one of the best predictors of turnover and that the literature findings show a negative relationship between tenure and turnover.

Compensation and benefits practices:

Several lines of evidence suggest that compensation and benefits practices have an important influence on voluntary turnover. Bennett *et al.* (1993), Getachew (2016), Getachew (2017) and Ayuure (2013) suggest that benefits, direct and indirect compensation contribute to the increase of

the opportunity cost of switching. In the same vein, Colding (2015), Negi (2013), Wang and Chen (2013), Arokiasamy (2013), Varadharaj and Irfan (2019), Workagegn (2017), Hussein (1989), Abeble (2016), Zhang *et al.* (2018), Selhadin (2019) and Tuji (2013) argue that satisfaction with pay and benefits is significant for retaining employees.

Communication:

According to Negassa (2016), employees have a strong need to be informed, and communication between the organization and its level determines employee job satisfaction. Factors such as the type of communication, feedback, sincerity, ethics, awareness, information asymmetry and respecting opinions can lead to employee turnover (Negassa, 2016; Parker and Skitmore, 2005; AlSayyed and Braiki, 2015).

Lack of flexibility:

Time flexibility has a positive effect on maintaining employment (Hochwarter *et al.*, 1999). Stress from overload and work-life imbalance can give rise to employee turnover (Abeble, 2016).

Lack of job security:

Type of national contract:

Recognition

Negassa (2016) points out that recognition involving prestige, opportunities, development, and recognition applied in the organization are avoidable workplace causes of employee attrition. Moreover, feeling unappreciated and to have a low social status and responsibility can affect turnover (Parker and Skitmore, 2005; Wang and Chen, 2013; Imani, 2013).

Organizational factors:

Size/Location:

In some geographical areas there may be more problems of various kinds related to the management of the subsidiaries.

2. Methodology

The methodological approach taken in this study is the Microsoft's TDSP framework, Team Data Science Process [17], also cited in Fallucchi *et al.* (2020). The TDSP is an agile, iterative data science process aimed to offer predictive analytics solutions and intelligent applications efficiently.
The TDSP approach was adopted to capture employee attrition motivations as well as to build a predictive model and consists in five phases:

2.1 Business understanding: the first step in this process was to specify the key variables that, according to the literature, allow to investigate the phenomenon of attrition. The previous section has shown the procedure followed for this purpose.
2.2 Data acquisition and understanding: the employee dataset was then collected and cleaned. The characteristics of the dataset have been described in the previous chapter, the variables

that are key to studying attrition were identified and the relative assumptions have been outlined.

2.3 Modeling: this phase aimed to describe the organization's population and to provide a variety of exploratory analysis that are relevant for the understanding of the key factors and trends that contribute to attrition (chapter 3).

2.4 Deployment: the prediction model was designed and implemented to identify employees that would potentially leave the company (chapter 4).

2.5 Customer acceptance: in the final stage of the study, the qualities of the adopted machine learning models were evaluated. The results have been shared with the organization and a revision of the models has been applied where it has emerged to be necessary (chapter 5).

3. Exploratory analyses

To begin this process, we perform a general observation of the data structure. Fig. 6 shows an overview of the dataset.

```
 CodicePersona          Genere               TitoloStudio  ClasseStruttura DettaglioStruttura
 Length:5767        Femmina:2050   Licenza elementare:  18   Centro :1848    DT01   : 707
 Class :character   Maschio:3717   Licenza media     :  62   Rete   :3886    DT07   : 595
 Mode  :character                  Diploma           :2480   Società:  33    DT09   : 556
                                   Laurea            :2807                   DT04   : 477
                                   Master            :  98                   DT03   : 334
                                   Altro             :  10                   DT08   : 330
                                   NA's              : 292                   (Other):2768
                                     Ruolo            Posizione         Eta         AnniServizio        Orario
 CA - CASSIERE                          : 390   Addetto     :4997   Min.   :19.00   Min.   : 0.00   Full Time:5460
 RDF - RESPONSABILE DI FILIALE          : 348   Responsabile: 770   1st Qu.:35.00   1st Qu.: 7.00   Part Time: 307
 CCE - CASSIERE COMMERCIALE EXPERT      : 343                       Median :45.00   Median :14.00
 CSA - CUSTOMER ASSISTANT               : 331                       Mean   :44.23   Mean   :15.82
 PBA - PERSONAL BANKER AFFLUENT         : 317                       3rd Qu.:54.00   3rd Qu.:25.00
 GSF - GESTORE SMALL BUSINESS DI FILIALE: 235                       Max.   :71.00   Max.   :44.00
 (Other)                                :3803
     Grado      QuotaOreLavoro               Contratto    Talento         Premio         RalMensile    IndennizzoRetention
 3.1    : 806   Min.   :0.0000   Credito           :5566   No  :5460   Min.   :    0   Min.   : 1235   Min.   :    0.00
 3.4    : 790   1st Qu.:1.0000   Dirigenti credito : 118   Sì  : 297   1st Qu.:  850   1st Qu.: 2954   1st Qu.:    0.00
 3.3    : 722   Median :1.0000   Portieri e custodi:   1   NA's:  10   Median : 2265   Median : 3697   Median :    0.00
 3.2    : 571   Mean   :0.9803   Altro             :  82               Mean   : 4980   Mean   : 4241   Mean   :   52.89
 4.2    : 560   3rd Qu.:1.0000                                        3rd Qu.: 5040   3rd Qu.: 4989   3rd Qu.:    0.00
 4.1    : 553   Max.   :1.0000                                        Max.   :500000   Max.   :54167   Max.   :50000.00
 (Other):1765   NA's   :82
   Gratifica          Citta            Indirizzo          Provincia          NumeroFigli      Potenziale    RetribuzioneAbi
 Min.   :    0.0   Length:5767        Length:5767        Length:5767        Min.   :0.000   Min.   : 0.00   Min.   : 2655
 1st Qu.:    0.0   Class :character   Class :character   Class :character   1st Qu.:0.000   1st Qu.:34.00   1st Qu.: 3226
 Median :    0.0   Mode  :character   Mode  :character   Mode  :character   Median :1.000   Median :39.00   Median : 3494
 Mean   :  179.1                                                           Mean   :1.077   Mean   :37.71   Mean   : 4267
 3rd Qu.:    0.0                                                           3rd Qu.:2.000   3rd Qu.:42.00   3rd Qu.: 5097
 Max.   :35000.0                                                           Max.   :7.000   Max.   :60.00   Max.   :20206
                                                                           NA's   :82      NA's   :933     NA's   :895

   Leadership        Clima       Dimissioni
 Min.   :4.550   Min.   :16.67   No:5458
 1st Qu.:8.031   1st Qu.:61.04   Sì: 309
 Median :8.600   Median :65.27
 Mean   :8.472   Mean   :65.33
 3rd Qu.:9.064   3rd Qu.:69.48
 Max.   :9.975   Max.   :95.49
 NA's   :5139    NA's   :177
```

Fig. 4: summary of the data

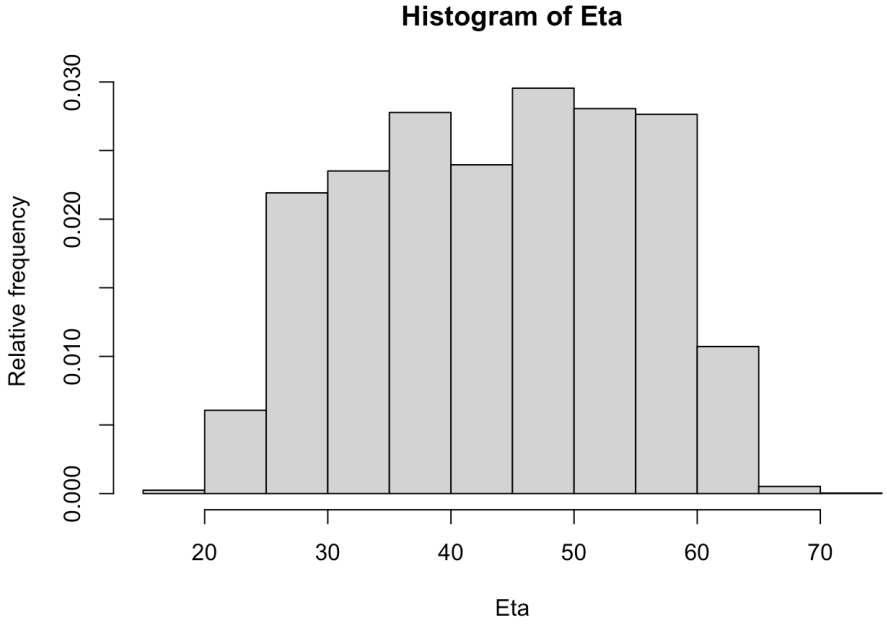
What follows are some explorative analyses that describe and visualize the organization's data.

3.1 Resignation by age


```
Dati$Dimissioni: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  19.00   36.00   45.00   44.64   54.00   71.00
------------------------------------------------------------
Dati$Dimissioni: Sì
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  21.00   29.00   35.00   37.03   44.00   63.00
```

Employees who resign are generally younger than those who do not.

The figure below illustrates the distribution of employees by age.



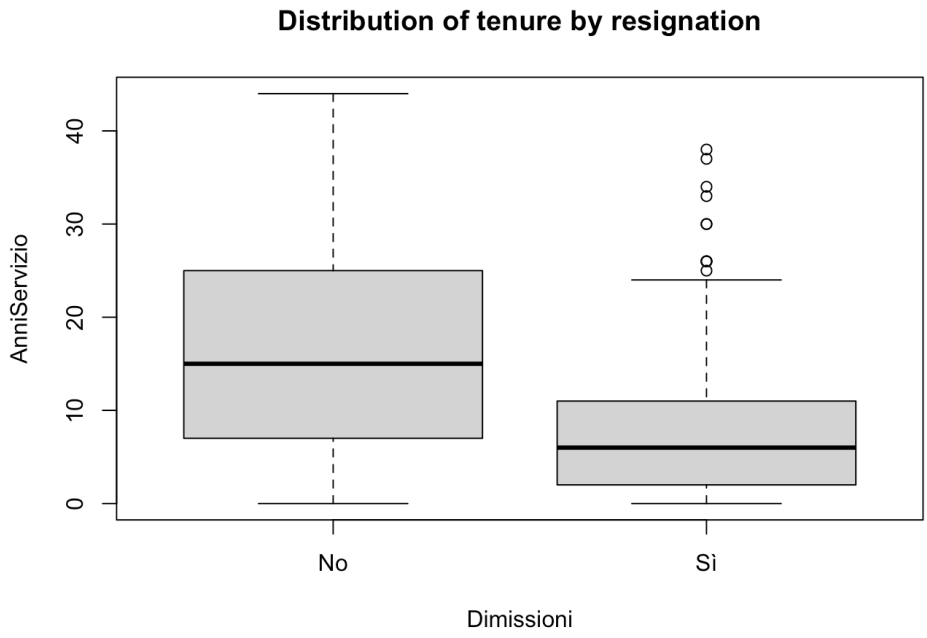
Devise a meaningful split of age into classes in order to assess the conditional frequency of resignations.
The relative frequency of resignations declines with age.

```
Resignation given age (%)
```

```
              No   Sì
  (0, 25]    84.1 15.9
  (25, 30]   90.0 10.0
  (30, 40]   94.4  5.6
  (40, 50]   97.0  3.0
  (50, +Inf] 98.6  1.4
```

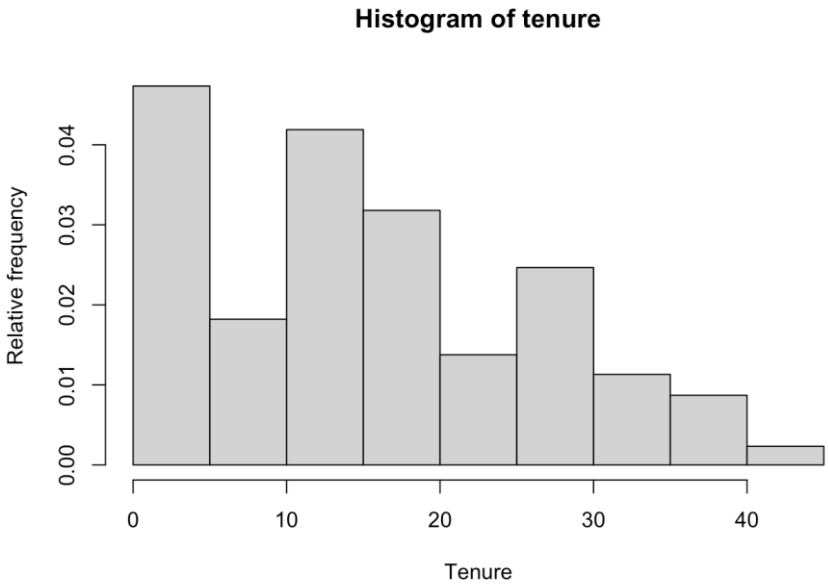
3.2 Resignation by tenure

Employees who resign generally have shorter tenure than those who do not.




```
Dati$Dimissioni: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    7.00   15.00   16.28   25.00   44.00
------------------------------------------------------------
Dati$Dimissioni: Sì
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   6.000   7.599  11.000  38.000
```

The figure below illustrates the distribution of employees by tenure.



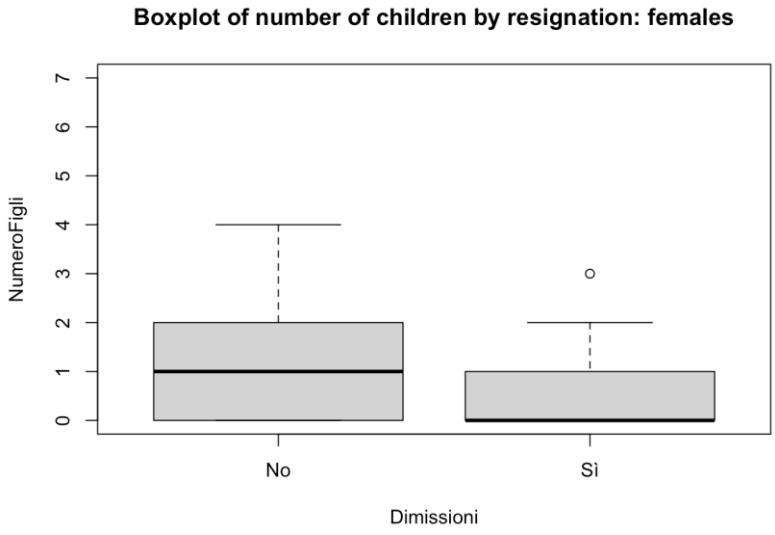
Devise a meaningful split of tenure into classes in order to assess the conditional frequency of resignations.
The relative frequency of resignations declines with tenure.

```
Resignation given tenure (%)
```

```
             No   Sì
(0, 3]     88.0 12.0
(3, 10]    89.4 10.6
(10, 20]   96.6  3.4
(20, 30]   98.6  1.4
(30, +Inf] 99.4  0.6
```

3.3 Resignation by gender and number of children

Employees who resign have less children than those who do not. The difference does not depend on gender.



Devise a meaningful split of number of children.

Resignations decrease with family care — for both genders.

```
Resignation given number of children (%)
```

```
             No   Sì
0          92.7  7.3
1          97.9  2.1
2          98.1  1.9
(2, 3]     98.5  1.5
(3, +Inf)  97.4  2.6
```

Then, are gender and resignation correlated?

```
Resignation by gender (%)
```

```
           No   Sì
Femmina 95.3  4.7
Maschio 94.3  5.7
```

Yes, slightly. Why? The following analyses show that interactions of features are likely to be significant.

3.3.1 Resignation by position

Resignation rate is higher in lower positions.

```
Reisgnation given position (%)
```

```
                No   Sì
Addetto       94.2  5.8
Responsabile  97.7  2.3
```

3.3.2 Position by gender

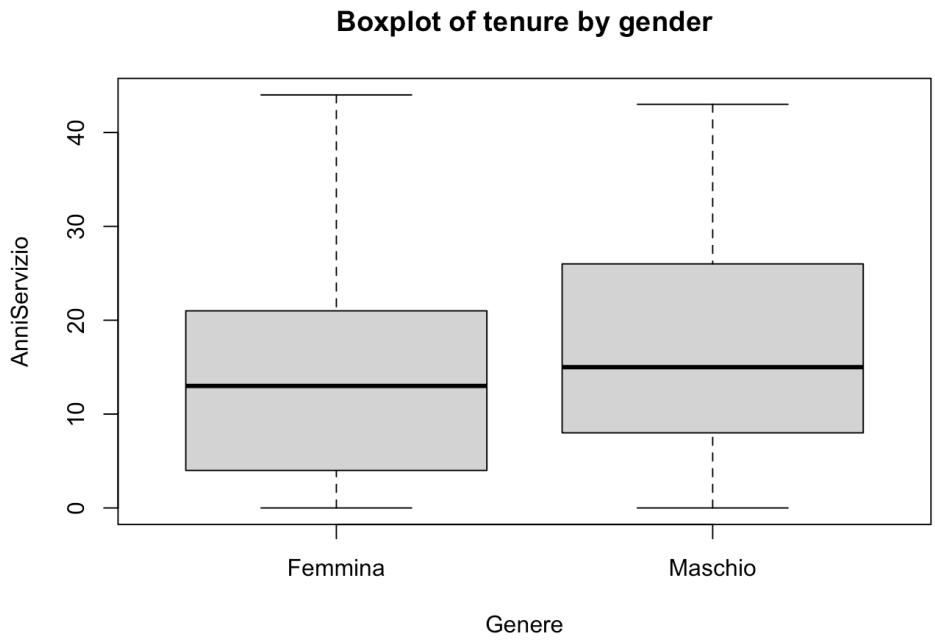
Higher positions are most frequent among male employees.

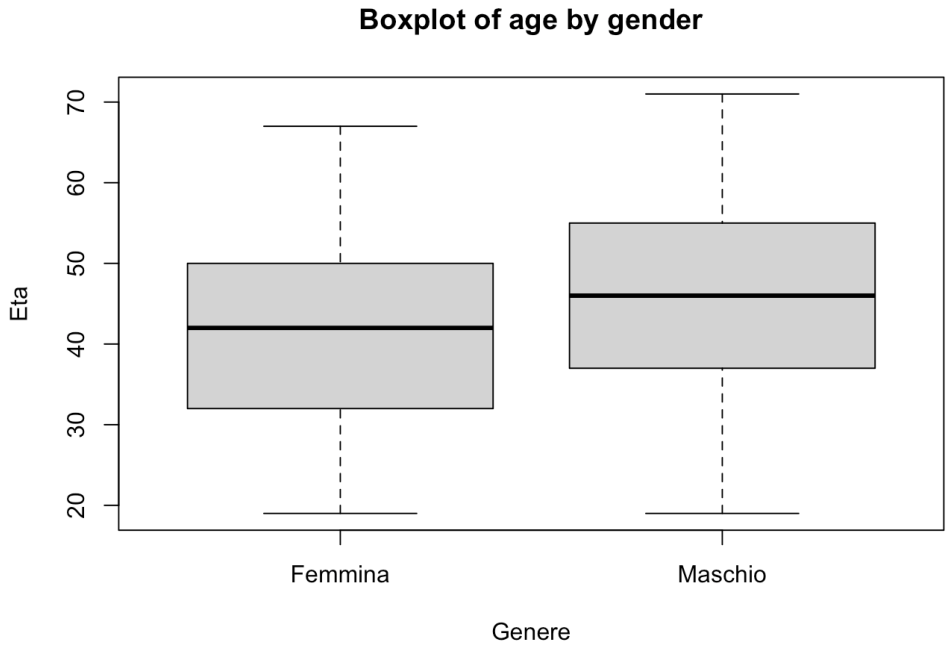
```
Gender given position (%)
```

```
         Addetto Responsabile
Femmina     93.3          6.7
Maschio     83.0         17.0
Sum         86.6         13.4
```

3.3.3 Tenure and age by gender

Males have longer tenures and are older than females.



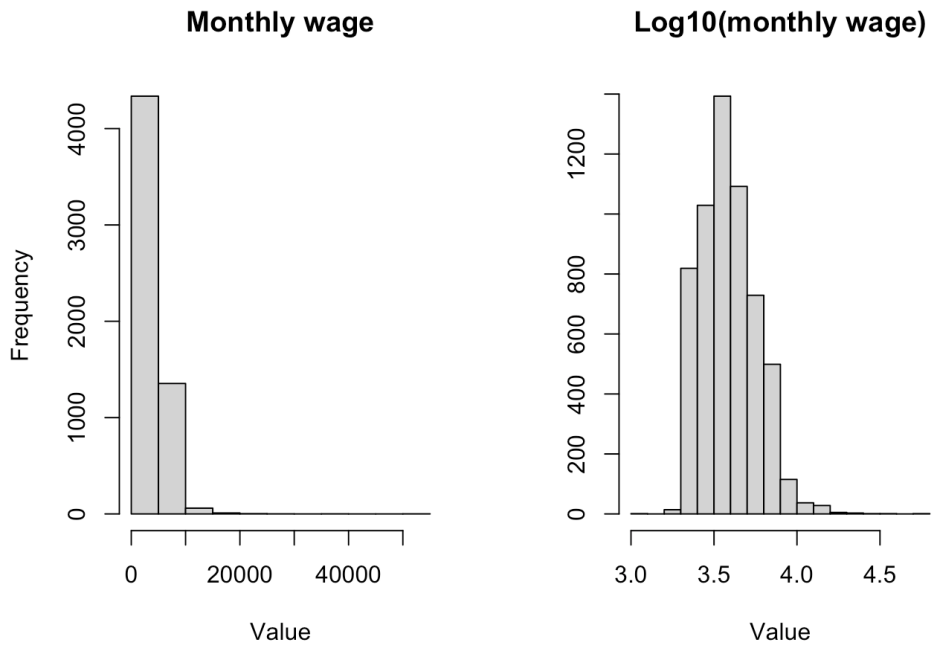


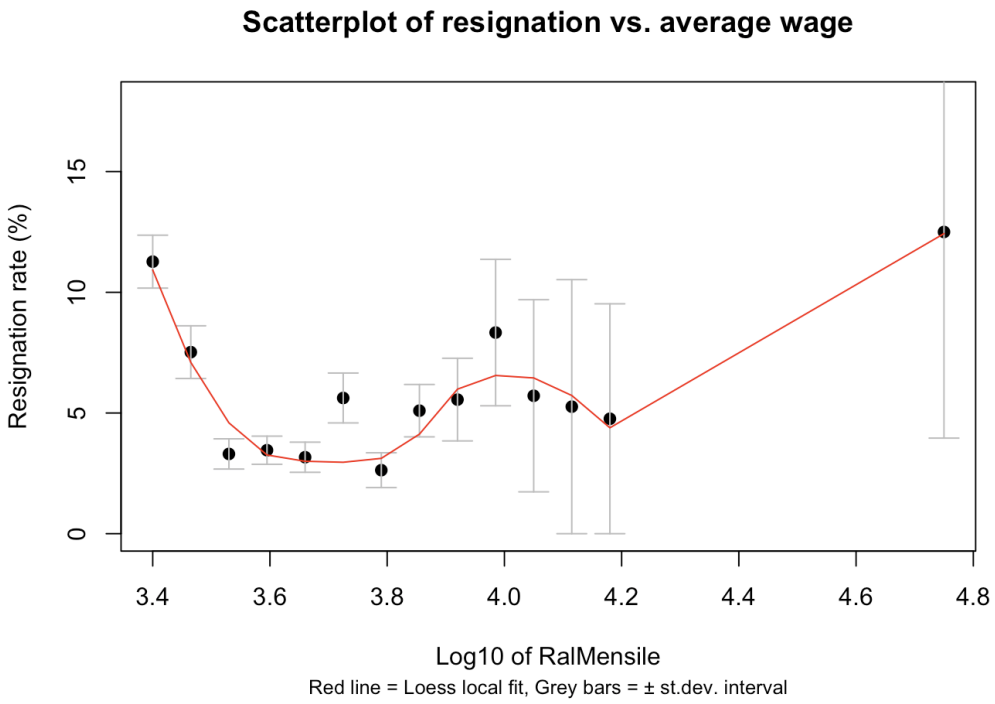
3.4 Resignation by wage

Because the distribution of wage is highly right-skewed, we analyse its logarithm (base 10).

The scatterplot shows that resignation rates depend on monthly wage in a non-linear fashion:

- decline with monthly for wage ≤ 3,600€;
- increase for 3,800 ≤ wage ≤ 4,000€;
- decline again for 4,000 < wage ≤ 4,200€.





The rightmost point in the plot represents top-level directors and is — naturally anomalous — with respect to the bulk of data.

Wage depends on several features that may, by themselves, affect resignation (e.g. job description, pay level, tenure).

3.4.1 Wage, job, pay grade

There are 209 job descriptions and 29 pay grades.
The distribution of (average) wage by job and pay grade are very dispersed.

Ruolo <fct>	RalMensileMedia <dbl>	N <int>
ADM - ASSET DEVELOPMENT MANAGER	4518	42
AMMMGT - AMMINISTRATIVO MGT	2955	1
APB - ASSISTANT PRIVATE BANKER	3375	72
ASB - ASSISTANT SMALL BUSINESS	3648	45
ASSALS - ASSISTANT AFFARI LEGALI E SOCIETARI	3914	13
ASSAMM - ASSISTANT AMMINISTRAZIONE	2769	2
ASSCOM - ASSISTANT COMPLIANCE E/O ANTIRICICLAGGIO	3573	6
ASSLOG - ASSISTANT IMMOBILIARE E LOGISTICA	2278	1
ASSORG - ASSISTANT ORGANIZZAZIONE	4267	2
ASSPEO - ASSISTANT PEOPLE MNGT	2982	3

1-10 of 209 rows Previous 1 2 3 4 5 6 … 21 Next

Grado <fct>	RalMensileMedia <dbl>	N <int>
2.1	2601	8
2.2	3039	3
2.3	2602	9
3.1	2481	806
3.2	2832	571
3.3	3109	722
3.4	3461	790
4.1	3912	553
4.2	4338	560
CDG	30000	1

1-10 of 29 rows Previous 1 2 3 Next

```
Average wage distribution by job
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   1235    3474    4679    5441    6313   30000 
```

```
Average wage distribution by pay grade
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   1235    3461    6890    9780   11026   45833 
```

Most variation in wage is by job and pay grade; interaction does not play any significant role.

```
Analysis of variance of log10(RalMensileMedia vs. job and pay grade
```

```
             Df Sum Sq Mean Sq
Ruolo       208 20.354  0.0979
Grado        25 14.311  0.5724
Ruolo:Grado 818  0.986  0.0012
```

3.4.2 Resignation by employment contract and pay grade

We have a problem with the “Contratto” feature, beacause the mode “Altro” belongs only to resigned employees. This is probably because resignation cases span several years in the past, while data on employees who did not resign a relative to 2019 only; something must have changed in the HRIS database in recent years.

Contratto <fct>	Grado <fct>	Rate <dbl>	StDev <dbl>	RalMensileMedia <dbl>	N <int>
Altro	2.3	100.0	0.0000	2146	3
Altro	3.1	100.0	0.0000	2278	14
Altro	3.2	100.0	0.0000	2623	10
Altro	3.3	100.0	0.0000	2725	7
Altro	3.4	100.0	0.0000	2773	3
Altro	4.1	100.0	0.0000	3642	10
Altro	4.2	100.0	0.0000	4368	8
Altro	COND	100.0	NA	8750	1
Altro	FUNZ2	100.0	0.0000	5384	4
Altro	FUNZ3	100.0	0.0000	4712	11

1-10 of 45 rows Previous 1 2 3 4 5 Next

Contratto	Dimissioni No	Dimissioni Sì
Credito	5344	222
Dirigenti credito	113	5
Portieri e custodi	1	0
Altro	0	82

To overcome the problem that we had an incomplete classification of the “Contratto” variable, that we consider relevant for our analysis, on a large number of persons who resigned (those referred to as "Altro”), we replaced these data with the corresponding ones through a manual intervention, since they were not in large numbers. This intervention was done by going to search in the dataset what were the correspondences between “Grado” and “Contratto”, and by replacing the variable “Altro” with the relative variable of “Contratto”. This substitution was made by intervening on the R code that was written to prepare the final dataset on which we carried out the analyses.

The table below illustrates the full values of the “Contratto” variable.

Contratto <chr>	Grado <fctr>	Rate <dbl>	StDev <dbl>	RalMensileMedia <dbl>	N <int>
Dirigenti credito	DG	50.0	50.0000	45833	2
Credito	2.3	33.3	16.6667	2602	9
Dirigenti credito	VDC	25.0	25.0000	15104	4
Credito	3.1	11.2	1.1101	2481	806
Dirigenti credito	VDP	7.7	7.6923	11026	13
Credito	PROC3	7.2	2.0044	6776	167
Credito	COND	6.7	4.6321	8464	30
Credito	VD1	6.4	3.6042	7868	47
Credito	FUNZ2	5.9	1.2743	5742	341
Credito	3.2	5.4	0.9491	2832	571
Credito	FUNZ3	5.4	0.9964	4924	517
Dirigenti credito	DD2	5.1	3.5782	9788	39
Credito	4.1	4.9	0.9172	3912	553

1–13 of 29 rows Previous 1 2 3 Next

```
                    Dimissioni
Contratto                 No   Sì
  Credito               5344  302
  Dirigenti credito      113    7
  Portieri e custodi       1    0
```

“Credito” is the prevailing contract and the one that contains most cases of resignation.

3.4.3 Gender and wage

Males have a higher wage than females.

Genere <fctr>	RalMensileMedia <dbl>	N <int>
Femmina	3556	2050
Maschio	4619	3717

3.4.4 Tenure and wage

As the tenure increases, wage increases.

ClasseAnniServizio <fctr>	RalMensileMedia <dbl>	N <int>
(0, 3]	3558	651
(3, 10]	4030	925
(10, 20]	4368	2125
(20, 30]	4583	1108
(30, +Inf]	4893	644
NA	2881	314

3.4.5 Age and wage

As the age increases, wage increases.

ClasseEta <fctr>	RalMensileMedia <dbl>	N <int>
(0, 25]	2447	182
(25, 30]	2837	1310
(30, 40]	3938	1492
(40, 50]	4996	1661
(50, +Inf]	5457	1122

3.4.6 Position and wage

The average wage in the highest position is almost twice the lower position.

Posizione <fctr>	RalMensileMedia <dbl>	N <int>
Addetto	3966	4997
Responsabile	6027	770

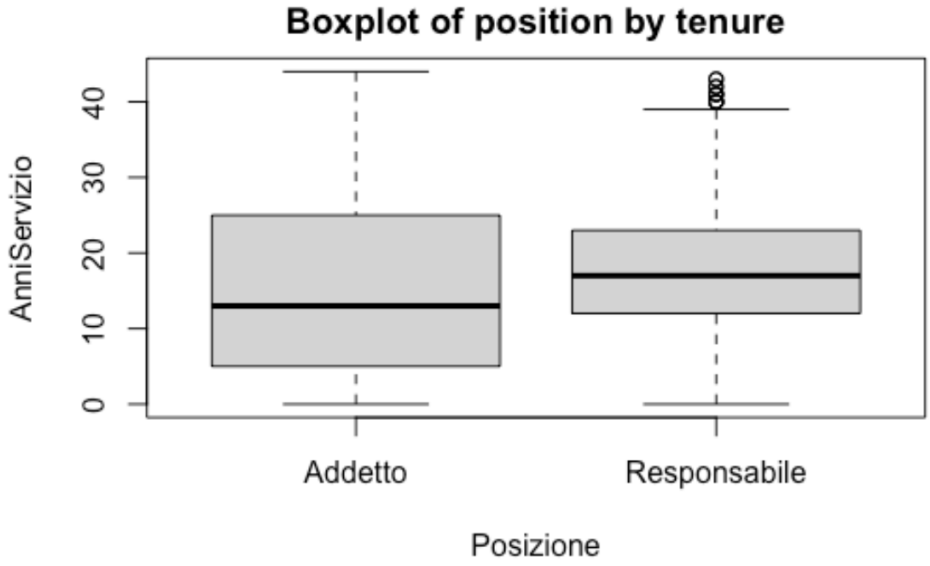
Results show that gender, tenure, age and position influence wage. We have previously investigated in detail the variables gender, tenure and age. For this reason, we are going to study the position variable in more detail.

3.4.7 Position and age

Employees in higher positions have an average age a little higher than the employees in the lower, but the difference is not significant.

3.4.8 Position and tenure

Employees in higher positions have an average tenure a little higher than the employees in the lower.



3.4.9 Position and talent

Higher positions are most frequent among talented employees.

Position by talent (%)

	No	Sì
Addetto	95.9	4.1
Responsabile	87.7	12.3

3.4.10 Position and educational level

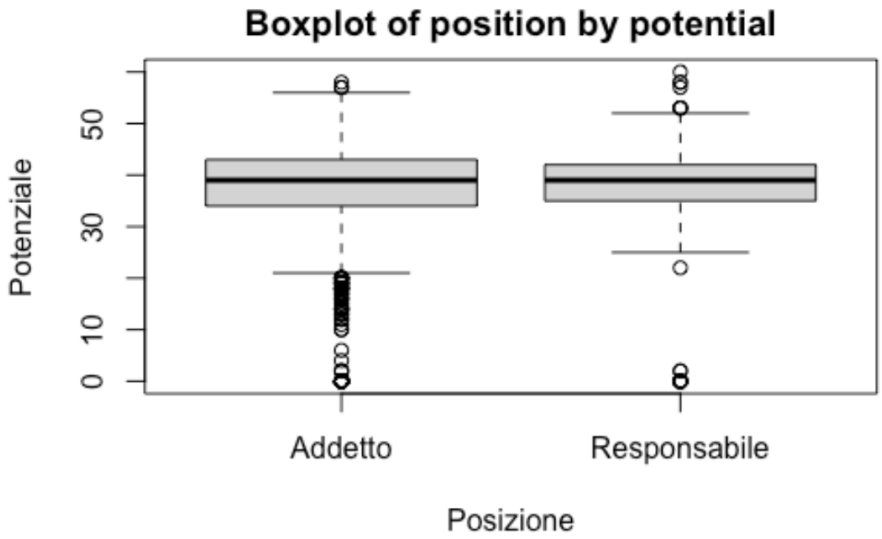
Most of the workers in higher positions holds a university degree (“Laurea”).

Position by educational level (%)

	Addetto	Responsabile
Licenza elementare	100.0	0.0
Licenza media	98.4	1.6
Diploma	87.9	12.1
Laurea	84.7	15.3
Master	79.6	20.4
Altro	70.0	30.0

3.4.11 Position and potential

The difference in position does not depend on potential.



3.4.12 Wage and % hours worked

(Average) wage increases with the percentage of hours worked.

ClasseQuotaOreLavoro <fctr>	RalMensileMedia <dbl>	N <int>
(0.25, 0.50]	2349	1
(0.50, 0.75]	2598	158
(0.75, 1]	4291	5491
NA	4125	117

3.4.13 Wage and working place

(Average) wage is a little higher in "Centro" than in the "Rete", while "Società" has higher (average) wage than both "Centro" and "Rete".

ClasseStruttura <fctr>	RalMensileMedia <dbl>	N <int>
Centro	4465	1848
Rete	4115	3886
Società	6564	33

3.5 Resignation by educational level

The relative frequency of resignations of the employees with the lowest levels of education is zero, while employees who hold a “Master” level degree have the highest percentage of resignation.

Resignation given educational level (%)

	No	Sì
Licenza elementare	100.0	0.0
Licenza media	100.0	0.0
Diploma	99.3	0.7
Laurea	99.5	0.5
Master	98.0	2.0
Altro	100.0	0.0
Sum	99.4	0.6

3.5.1 Gender and educational level

Females rate initially grows as the educational level grows, and, after university graduation ("Laurea") tends to decrease.

Gender given educational level (%)

	Femmina	Maschio
Licenza elementare	22.2	77.8
Licenza media	22.6	77.4
Diploma	31.6	68.4
Laurea	39.8	60.2
Master	36.7	63.3
Altro	30.0	70.0
Sum	35.8	64.2

3.5.2 Working place and educational level

As the educational level grows, the percentage of employees working in "Centro" increases, and in "Rete" decreases.

Working place given educational level (%)

	Centro	Rete	Società
Licenza elementare	16.7	83.3	0.0
Licenza media	25.8	74.2	0.0
Diploma	28.3	70.9	0.7
Laurea	35.3	64.3	0.4
Master	36.7	61.2	2.0
Altro	50.0	50.0	0.0
Sum	32.0	67.4	0.6

In “Centro” university graduates are mostly, in “Rete” university graduates are slightly in prevalence, followed by high-school graduates, and in “Società” the high-school graduates represent the majority. The highest percentage of people who own a Master is present in the “Società”.

Working place given educational level (%)

	Licenza elementare	Licenza media	Diploma	Laurea	Master	Altro
Centro	0.2	0.9	40.1	56.5	2.1	0.3
Rete	0.4	1.2	47.7	48.9	1.6	0.1
Società	0.0	0.0	58.1	35.5	6.5	0.0
Sum	0.3	1.1	45.3	51.3	1.8	0.2

3.5.3 Age and educational level

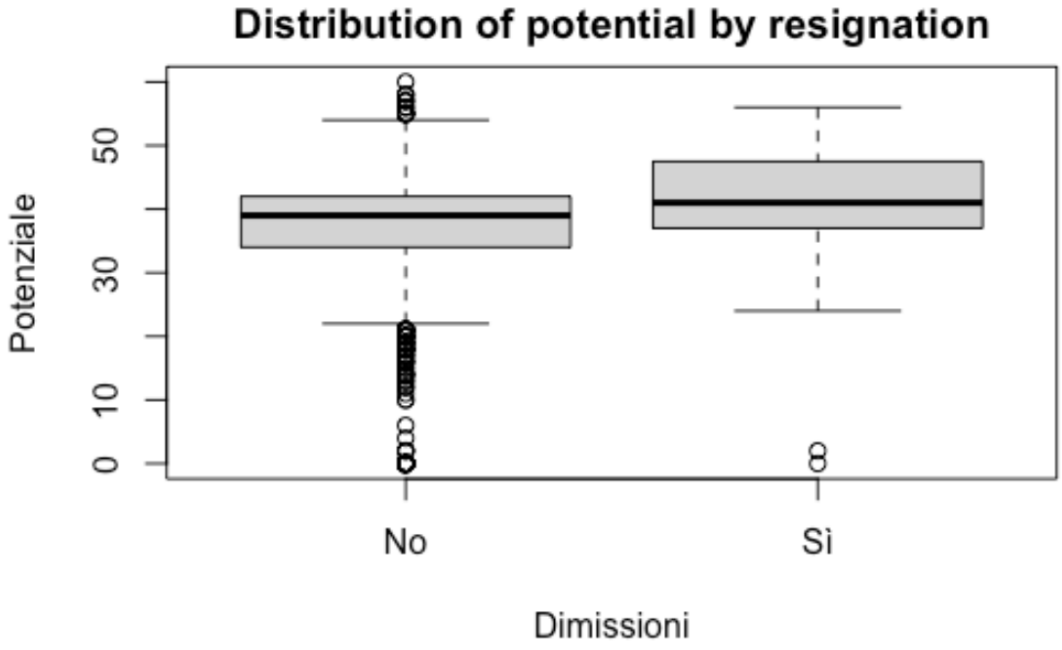
Younger people (up to 40 years of age) have on average higher levels of education.

Age given educational level (%)

	(0, 25]	(25, 30]	(30, 40]	(40, 50]	(50, +Inf]
Licenza elementare	0.0	0.0	5.6	50.0	44.4
Licenza media	0.0	0.0	0.0	53.2	46.8
Diploma	3.1	10.9	18.5	35.6	32.0
Laurea	2.5	31.9	31.8	23.9	9.8
Master	2.0	19.4	53.1	22.4	3.1
Altro	0.0	30.0	60.0	10.0	0.0
Sum	2.7	21.7	25.8	29.6	20.3

3.6 Resignation by potential and talent

Employees who resign generally have a higher potential and talent than those who do not.

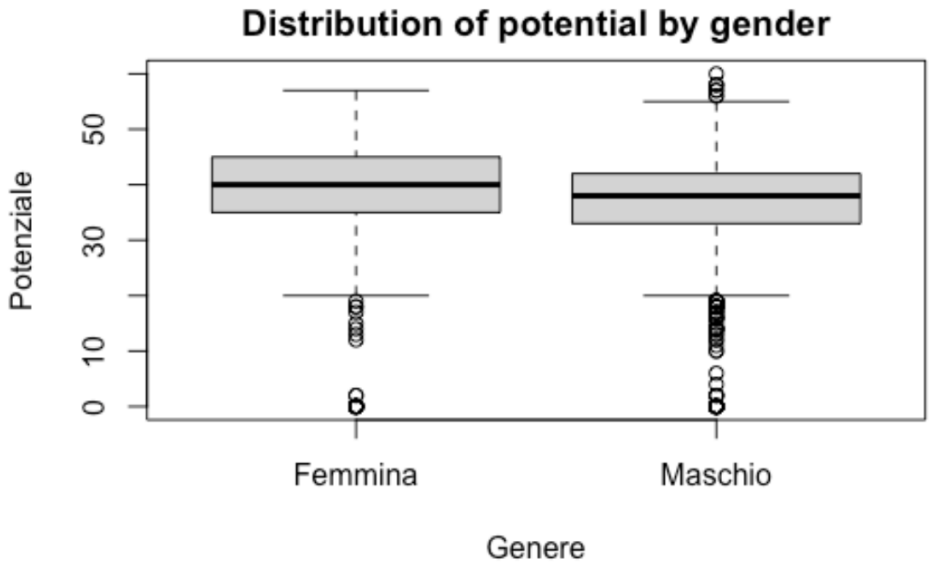


```
Dati$Dimissioni: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   34.00   39.00   37.68   42.00   60.00     676
------------------------------------------------------
Dati$Dimissioni: Sì
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   37.00   41.00   40.35   47.25   56.00     257

 Resignation given talent (%)
        No   Sì
   No  94.9  5.1
   Sì  93.3  6.7
   Sum 94.8  5.2
```

3.6.1 Gender and potential and talent

Female employees generally have a slightly higher potential than male employees. The percentage of talented women employees is a little higher than that of men employees.




```
Dati$Genere: Femmina
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   35.00   40.00   38.99   45.00   57.00     347
------------------------------------------------------
Dati$Genere: Maschio
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   33.00   38.00   37.02   42.00   60.00     586

Gender given talent (%)
            No   Sì
  Femmina 94.3  5.7
  Maschio 95.1  4.9
```

3.7 Resignation by % hours worked

The relative frequency of resignation increases with the percentage of hours worked.

```
 Resignation given % hours worked (%)
                    No    Sì
   (0, 0.50]    100.0   0.0
   (0.50, 0.95]  97.7   2.3
   (0.95, 1]     96.5   3.5


Dati$Dimissioni: No
   (0, 0.50] (0.50, 0.95]    (0.95, 1]
           1          295         5162
------------------------------------------------------
Dati$Dimissioni: Sì
   (0, 0.50] (0.50, 0.95]    (0.95, 1]         NA's
           0            7          185          117
```

We recalculated the values by redefining part of the class breakdown, (0.50, 0.90] and (0.90, 1], and we saw that the results do not change.

```
 Resignation given % hours worked (%)
                    No    Sì
   (0, 0.50]    100.0   0.0
   (0.50, 0.90]  97.7   2.3
   (0.90, 1]     96.5   3.5

Dati$Dimissioni: No
   (0, 0.50] (0.50, 0.90]    (0.90, 1]
           1          295         5162
------------------------------------------------------
Dati$Dimissioni: Sì
   (0, 0.50] (0.50, 0.90]    (0.90, 1]         NA's
           0            7          185          117
```

3.7.1 % hours worked and gender

It seems that women tend to favour working conditions that are not full-time. Women are most frequent in the (0.50, 0.95] class of % hours worked, while men are more frequent in the (0.95, 1] class.

```
% hours worked given gender (%)
               Femmina Maschio
  (0, 0.50]        0.0   100.0
  (0.50, 0.95]    93.4     6.6
  (0.95, 1]       32.5    67.5

Dati$Genere: Femmina
   (0, 0.50] (0.50, 0.95]    (0.95, 1]         NA's
           0           282         1737           31
------------------------------------------------------
Dati$Genere: Maschio
   (0, 0.50] (0.50, 0.95]    (0.95, 1]         NA's
           1            20         3610           86
```

The segmentation of class (0, 0.50] is too significant, because there is only one worker. We then redefined the classes and recalculated the values, which are reported in the following figures.

```
% hours worked given gender (%)
               Femmina Maschio
  (0, 0.75]       89.3    10.7
  (0.75, 0.90]    97.2     2.8
  (0.90, 1]       32.5    67.5

Dati$Genere: Femmina
   (0, 0.75] (0.75, 0.90]    (0.90, 1]         NA's
         142           140         1737           31
------------------------------------------------------
Dati$Genere: Maschio
   (0, 0.75] (0.75, 0.90]    (0.90, 1]         NA's
          17             4         3610           86
```

The last class remains the prevailing one. Women remains most frequent in the second class of % hours worked, while men are more frequent in the (0.90, 1] class.

3.8 Resignation by working place

Resignation rate in lower in "Centro" and higher in "Società", although the difference is not substantial.

```
Resignation given working place (%)
            No   Sì
  Centro  94.8  5.2
  Rete    94.6  5.4
  Società 93.9  6.1
```

3.8.1 % hours worked and working place

In “Rete” and in “Società” the proportion of full-time workers is greater.

```
% hours worked given working place (%)
               Centro  Rete Società
  (0, 0.50]     100.0   0.0     0.0
  (0.50, 0.95]   34.1  65.6     0.3
  (0.95, 1]      32.1  67.4     0.6

Dati$ClasseStruttura: Centro
   (0, 0.50] (0.50, 0.95]    (0.95, 1]         NA's
           1          103         1715           29
------------------------------------------------------
Dati$ClasseStruttura: Rete
   (0, 0.50] (0.50, 0.95]    (0.95, 1]         NA's
           0          198         3602           86
------------------------------------------------------
Dati$ClasseStruttura: Società
   (0, 0.50] (0.50, 0.95]    (0.95, 1]         NA's
           0            1           30            2
```

Even in this case, the segmentation of class (0, 0.50] is too significant, because there is only one worker. We then redefined the classes and recalculated the values, which are reported in the following figures.

```
% hours worked given working place (%)
               Centro Rete Società
  (0, 0.75]      30.2 69.8     0.0
  (0.75, 0.90]   38.9 60.4     0.7
  (0.90, 1]      32.1 67.4     0.6

Dati$ClasseStruttura: Centro
   (0, 0.75] (0.75, 0.90]    (0.90, 1]         NA's
          48           56         1715           29
------------------------------------------------------
Dati$ClasseStruttura: Rete
   (0, 0.75] (0.75, 0.90]    (0.90, 1]         NA's
         111           87         3602           86
------------------------------------------------------
Dati$ClasseStruttura: Società
   (0, 0.75] (0.75, 0.90]    (0.90, 1]         NA's
           0            1           30            2
```

The last class remains the prevailing one.

3.8.2 Gender and working place

There is no great gender difference in the three different working places. The percentage of women is a little higher in "Centro", while the percentage of men in "Rete".

```
Gender given working place (%)
          Centro Rete Società
  Femmina   34.0 65.5     0.5
  Maschio   30.9 68.4     0.6
```

3.9 Resignation by part-time/full-time schedule

Full-time employees have a higher percentage of resignation.

```
Resignation given part-time/full-time (%)
              No   Sì
  Full Time 94.5  5.5
  Part Time 96.4  3.6
```

3.9.1 Gender and part-time/full-time schedule

Full-time job is mostly covered by male employees, while part-time by women employees.

```
Gender given part-time/full-time (%)
            Femmina Maschio
  Full Time    32.3    67.7
  Part Time    93.2     6.8
```

3.9.2 Working place and part-time/full-time schedule

Part-time job is in the minority in all three working places. In "Centro" is a little higher than in "Rete" and "Società".

```
Working place given part-time/full-time (%)
          Full Time Part Time
  Centro       94.3       5.7
  Rete         94.8       5.2
  Società      97.0       3.0
```

3.9.3 Educational level and part-time/full-time schedule

The proportion of part-time and full-time workers does not vary much with the educational level.

```
Educational level given part-time/full-time (%)
                     Full Time Part Time
  Licenza elementare      94.4       5.6
  Licenza media           95.2       4.8
  Diploma                 93.5       6.5
  Laurea                  95.5       4.5
  Master                  94.9       5.1
  Altro                  100.0       0.0
```

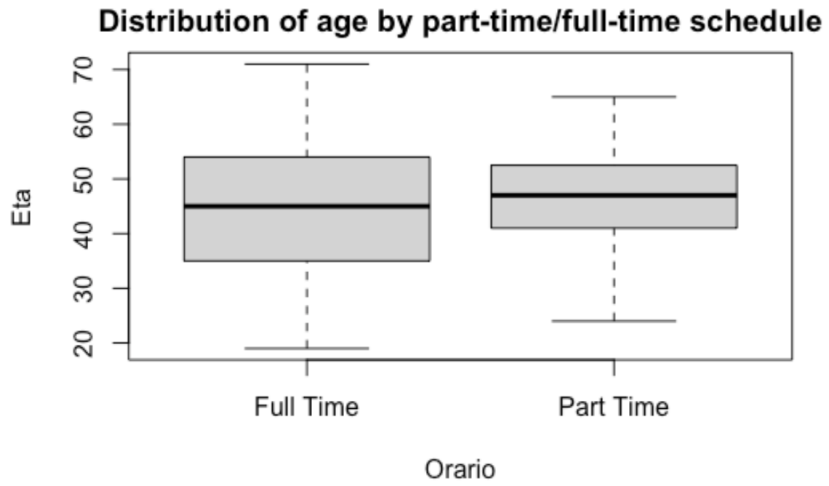
3.9.4 Position and part-time/full-time schedule

Part-time job rate is higher in lower positions.

```
 Position given part-time/full-time (%)
               Full Time Part Time
   Addetto           93.9       6.1
   Responsabile      99.9       0.1
```

3.9.5 Age and part-time/full-time schedule

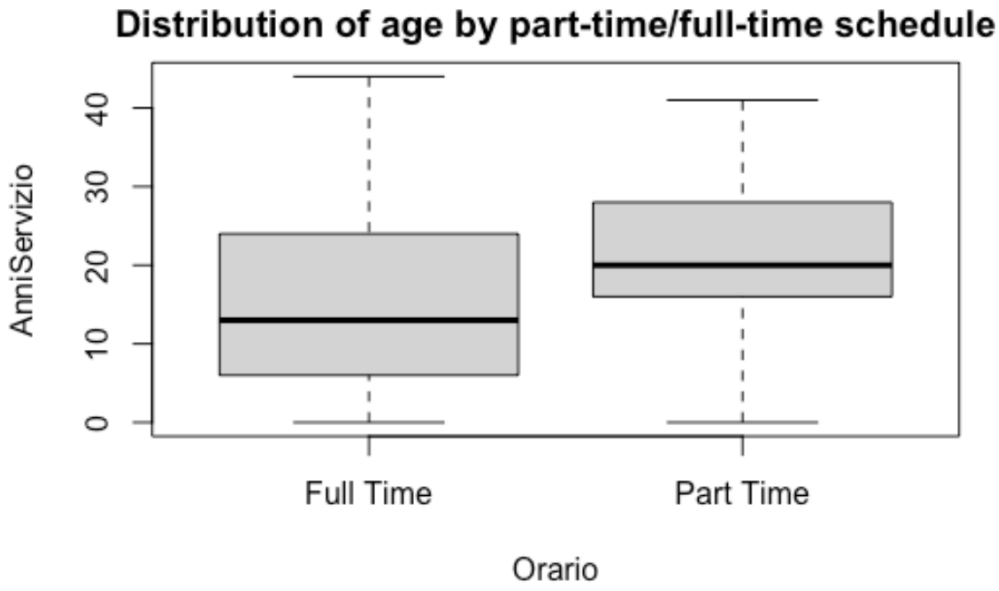
Part-time employees are generally a little older than full-time employees.




```
Dati$Orario: Full Time
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  19.00   35.00   45.00   44.07   54.00   71.00
------------------------------------------------------
Dati$Orario: Part Time
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  24.00   41.00   47.00   47.09   52.50   65.00
```

3.9.6 Organizational tenure and part-time/full-time schedule

Employees who have a part-time job generally have higher tenure than those who not.




```
Dati$Orario: Full Time
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    6.00   13.00   15.51   24.00   44.00
------------------------------------------------------
Dati$Orario: Part Time
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   16.00   20.00   21.29   28.00   41.00
```

Taken together, these results suggest that part-time choice would appear to be associated with less career-interested profiles.

4. Prediction model

4.1 Techniques to predict employee attrition: an overview

In the context of the employee attrition investigations, models and machine learning tools are designed, created and applied for three different goals: (I) predicting employee attrition and (II) identify how many of the churned employees were “valuable”; (iii) overcome class imbalance.

We adopted a part of the classification used by Fareri *et al.* (2020) to summarize the three different goals and the models, algorithms and machine learning techniques suggested for achieving them (Fig. 1).

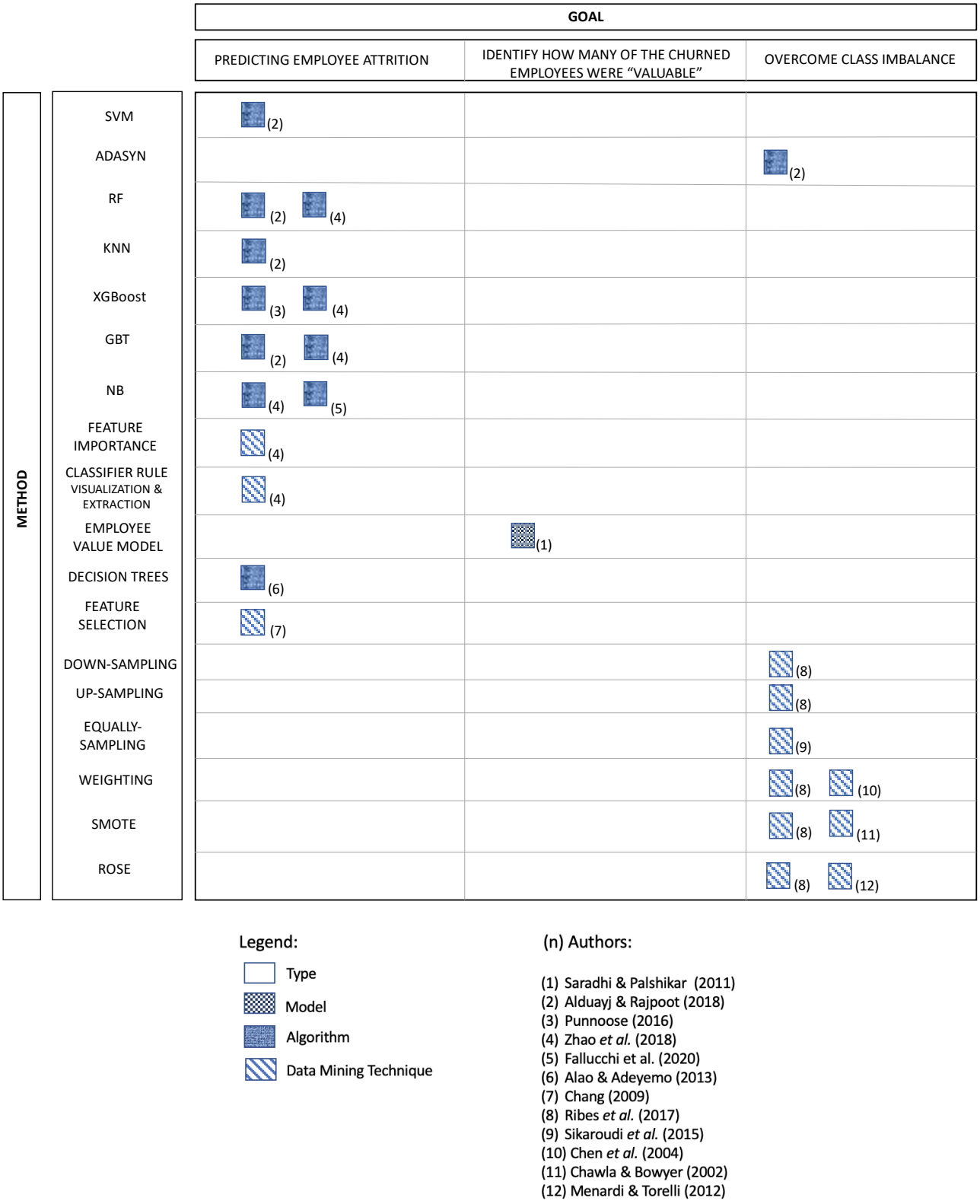


Figure 4: models and machine learning tools literature map. The map should be read as follows: the author "n" tries to achieve the goal "k", using the "p" type of tool.

The main methods suggested by the literature with the aim of achieving the objectives mentioned above are:

- **Employee value model:** a model used to identify the valuable employee among those predicted to churn (Saradhi and Palshikar, 2011).
- **Support Vector Machines (SVMs):** a non-probabilistic supervised machine learning model used for classification and regression (Alduayj and Rajpoot, 2018). Quadratic SVM scored the highest results (F1 score) in training the original class-imbalanced dataset (Alduayj and Rajpoot, 2018).
- **Random Forest (RF):** a supervised machine learning algorithm for generating classifications and regressions that uses multiple decision trees to train data (Alduayj and Rajpoot, 2018). In the study of Zhao *et al.* (2018), it ranked third, after the XGBoost and the GBT. It is an extension of decision trees and provided a marginal increase of performance at the addition of increased complexity in the study of El-Rayes *et al.* (2020).
- **K-nearest neighbours (KNN):** a machine learning algorithm used for classification and regression that works by specifying the values of K (Alduayj and Rajpoot, 2018).
- **Extreme Gradient Boosting (XGBoost):** a boosted tree algorithm that follows the principle of gradient boosting. It is very useful to control over-fitting (Punnoose, 2016). The study by Punnose (2016) explored the application of XGBoost technique in predicting employee turnover using noise-ridden data from the HRIS of a global retailer. He compared XGBoost classifier against six other supervised classifiers that had been historically used to build turnover models. He demonstrated that XGBoost reported significantly higher accuracy for predicting turnover and proved to be capable of handling the noise in the data from HRIS. Similarly, in the assessment of Zhao *et al.* (2018) of supervised machine learning methods for handling imbalanced HR datasets which contains noise and missing values, XGBoost had the best overall performance.
- **Gradient Boosting Trees (GBT):** a machine learning method for regression and classification in which the gradient boosted tree models learn sequentially (Zhao *et al.*, 2018). In the study of Zhao *et al.* (2018), it ranked second, after the XGBoost, and performed best for the bank datasets.
- **Naïve Bayes (NB):** a probabilistic approach that describes the occurrence probability of an event based on the prior knowledge of related features (Zhao *et al.*, 2018). In the study of Fallucchi et al. (2020), Gaussian Naïve Bayes was identified as the best algorithm in predicting the greatest number of people who could leave the company by minimising the number of false negatives.
- **Decision Trees:** a tree-shaped structures that represent decision sets. Its goal is to create a model that predicts the value of a target variable based on several input variables (Alao and Adeyemo, 2013).

  The most common types of decision tree algorithm are:

  -**CHAID** (Chi-square automatic interaction detection): it can produce tree with multiple sub-nodes for each split (Alao and Adeyemo, 2013);

  -**C4.5:** based on information theory, is an extension of Quinlan's earlier ID3 algorithm (Alao and Adeyemo, 2013). It builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy (Alao and Adeyemo, 2013).

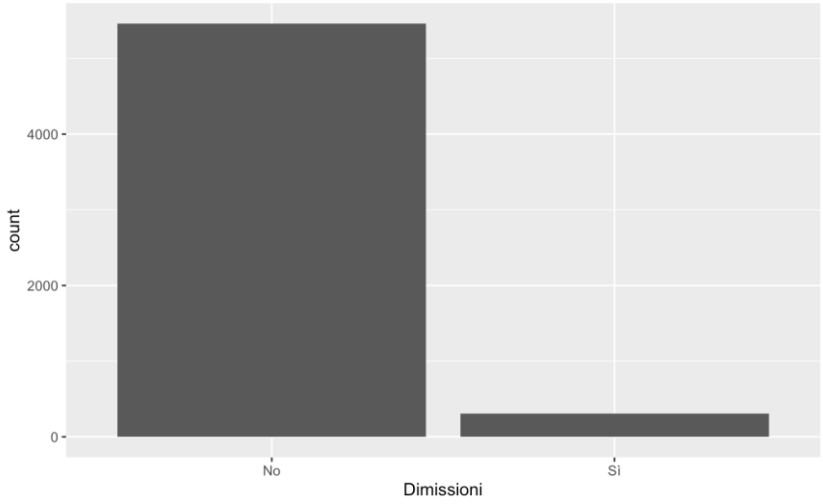
  -**CART (Classification and Regression Tree):** recursively partitions on a nominal target category to reach a tree structure (Sikaroudi et al., 2015)

In their analysis, El-Rayes *et al.* (2020) found that random forest and decision tree methods were the strongest attrition prediction models. In particular, decision trees provide the strongest predictive performance relative to model complexity.

- **Feature Importance:** a data mining technique useful for determine the influence of specific feature in affecting employee turnover as a whole and understanding the correlations between features and employee turnover. It should be calculated after identifying the best performance classifier (Zhao *et al.*, 2018). Most tree-based classifiers fall into this category (Zhao *et al.*, 2018).
- **Feature Selection:** a process that select important or relevant feature from the original feature set and offers many advantages for pattern classification (Chang, 2009). The idea is closely related to Feature Importance (Zhao *et al.*, 2018).
- **Classifier Rule Visualization and Extraction:** a method to convert machine learning models into easy-to-understand, interpretable figures or sets of rules (Zhao *et al.*, 2018).

Class Imbalance Correction Methods

Class Imbalance is a common theme on employee churn prediction. As Chawla *et al.* (2002) observe: "A dataset is imbalanced if the classification categories are not approximately equally represented". This is our case: out of the total 5767 observations, 309 have resigned, compared to 5458 that have stayed.



	Dimissioni	freq
1	No	5458
2	Sì	309

Fig. 5: count of attrition

In this case, the problem lies in the fact that standard classifiers tend to be overwhelmed by the dominant class and ignore the rare examples (Menardi & Torelli, 2012). The following are some of the main methods proposed in the literature to solve this problem:

- **Down-sampling:** sample the majority class to make its frequency closer to the rarest class (Ribes *et al.*, 2017; Sikaroudi *et al.*, 2015).

- **Up-sampling:** resample the minority class to increase its frequency (Ribes *et al.*, 2017). It yielded the best performance, together with ROSE, in the experiment of Ribes *et al.* (2017).
- **Weighting:** place a heavier penalty on misclassifying the minority class**.** It assigns a weight to each class, with the minority class given larger weight (i.e., higher misclassification cost) (Ribes *et al.*, 2017; Chen *et al.*, 2004).
- **"SMOTE" (Synthetic Minority Over-sampling Technique):** over-sample the minority class by creating "synthetic" examples rather than by over-sampling with replacement. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors (Ribes *et al.*, 2017; Chawla *et al.*, 2002). Neighbors from the k nearest neighbors are randomly chosen. It can be combined with under-sampling the majority class by randomly removing samples from the majority class, until the minority class becomes some specified percentage of the majority class (Ribes *et al.*, 2017; Chawla *et al.*, 2002). In the study conducted by Chawla *et al.* (2002), it was shown that the SMOTE approach can improve the accuracy of classifiers for a minority class. They also found that the combination of SMOTE and under-sampling also performs better, based on domination in the ROC space.
- **"ROSE":** generate new artificial data from the classes, according to a smoothed bootstrap approach, in order to address the same attention to both the classes. It combines technique to oversampling and undersampling by generating an augmented sample of data, especially belonging to the rare class (Ribes *et al.*, 2017; Menardi & Torelli, 2012). It yielded the best performance, together with up-sampling, in the experiment of Ribes *et al.* (2017).
- **Adaptive Synthetic (ADASYN) sampling approach:** an algorithm that solves the class imbalance by creating new synthetic instances based on the density distribution of the minority class (Alduayj and Rajpoot, 2018). In the experiment conducted by Alduayj and Rajpoot (2018), using ADASYN approach to overcome class imbalance, Cubic SVM, Gaussian SVM, RF and KNN achieved the highest results (F1 score).

The table below illustrates explains in detail the predictive models that have been analyzed by the authors, which performance measures were used to evaluate them and the results they arrive at (Table 3).

Table 3: literature framework with details of the models analysed, the performance measurements used and the result

Authors	Topic	Evalutation Method	Models		Values of Criteria %									Recommended
					AUC on Train	AUC on Test	Accuracy on k-fold validation	Accuracy on Train	Accuracy on Test	Precision on Test	F1 score on Test	True Positive Rate on Test	True Negative Rate on Test	
Sharadhi & Palshikar (2011)	Conduct a case study for building and comparing predictive employee churn models and propose a value model for employees	Three datasets: (1) resigned and released, (2) only resigned, (3) only released. 80:20 random split into train and test sets	SVM (1)						80,0			81,2	79,7	SVM
			SVM (2)						82,0			80,0	82,2	
			SVM (3)						84,1			77,8	85,1	
			Random forests (1)						83,5			53,6	91,9	
			Random forests (2)						92,3			50,0	97,5	
			Random forests (3)						88,8			47,2	95,0	
			Naïve Bayes (1)						74,0			55,1	79,3	
			Naïve Bayes (2)						79,0			63,3	81,0	
			Naïve Bayes (3)						77,6			44,4	82,6	
Alduayi & Rajpoot (2018)	Conduct three main experiments to predict employee attrition using unbalanced and synthetic data from IBM Watson	On an unspecified split of trin and test set	Unbalanced data	Linear SVM					86,9	81,4	37,1	24,0		Quadratic SVM
				Quadratic SVM					87,1	66,2	50,3	40,5		
				Cubic SVM					84,1	50,8	45,8	41,8		
				Gaussian SVM					86,5	78,8	34,3	21,9		
				Random Forest					85,6	75,0	26,9	16,4		
				k Nearest neighbors (K=1)					82,7	27,5	7,9	4,6		
				k Nearest neighbors (K=3)					83,7	25,0	0,8	0,4		
			Synthetic data	Linear SVM					78,2	76,3	77,9	79,5		Cubic, Gaussian, random forest and k Nearest neighbors
				Quadratic SVM					87,9	83,9	88,1	92,7		
				Cubic SVM					92,6	87,9	92,7	98,1		
				Gaussian SVM					91,2	88,5	91,2	94,1		
				Random Forest					92,6	95,0	92,1	89,3		
				k Nearest neighbors (K=1)					96,7	93,9	96,7	99,7		
				k Nearest neighbors (K=3)					92,9	87,7	93,1	99,2		
				k Nearest neighbors (K=5)					90,4	84,3	90,9	98,7		
				k Nearest neighbors (K=10)					87,2	80,4	88,0	97,0		

Punnose (2016)	Compare XGBoost technique against six historically used classifier using data from HRIS of a global retailer	Dataset split 80:20 into train and test set	Random forest (XGBoost)	88,0	86,0							XGBoost
			Logistic Regression	66,0	50,0							
			Naïve Bayes	64,0	59,0							
			Random Forest (Depth controlled)	79,0	51,0							
			SVM (Radial Basis Function)	68,0	52,0							
			Linear discriminant analysis	74,0	52,0							
			k Nearest neighbors	52,0	50,0							
Fallucchi *et al.* (2020)	Analyse a real dataset provided by IBM analytics and identify the main causes that influence attrition	Dataset split 70:30 into train and test set	Naïve Bayes (Gaussian)			78,2	82,5	38,6	44,6	54,1	84,5	Naïve Bayes (Gaussian)
			Naïve Bayes (Bernoulli)			83,1	84,5	45,9	37,9	33,1	92,7	
			Logistic Regression			86,5	87,5	66,3	44,5	33,7	96,2	
			k Nearest Neighbour			84,2	85,2	55,1	15,0	9,0	99,4	
			Decision Tree			79,2	82,3	35,6	35,1	36,1	91,0	
			Random Forest			85,0	86,1	65,8	19,4	13,2	99,1	
			SVM			85,1	85,9	80,8	16,6	9,6	99,4	
			Linear SVM			85,8	87,9	66,5	35,8	24,7	97,8	
Alao & Adeyemo (2013)	Analyse complete records of employees of one of the Higher Institutions in Nigeria, identifying employee related attributes that contribute to the prediction of attrition	On an unspecified split of trin and test set	CART (C4.5)		78,4			61,3	63,6	67,0	85,7	CART (C4.5)
			CART (REPTree)		74,9			55,3	57,9	61,8	84,4	
			CART (Basic)		77,7			57,9	60,8	64,1	86,3	
El-Rayes *et al.* (2020)	Develop tree-based binary classification models to predict the likelihood of employee attrition based on firm cultural and management attributes	Dataset split randomly 80:20 into train and test set	Linear regression	65,0								Random forest and decision tree
			Logistic regression	65,0								
			Decision Tree	70,0								
			Random Forest	73,0								
Ribes *et al.* (2017)	Illustrating the similarities between the problem of customer churn and employee turnover and developing an example of employee turnover prediction model	Dataset split 80:20 into train and test set	Linear discriminant analysis		75,0					74,0	33,0	Tree based ones
			SVM (Radial Basis Function)		80,0					76,0	31,0	
			Random Forest		95,0					96,0	19,0	
			Tree bagging		94,0					94,0	22,0	
Sikaroudi *et al.* (2015)	Applying different data mining methods on real data of manufacturing plant to predict employee turnover.	k-fold cross validation	Multy-layer perceptron			82,0						Decision trees
			Probabilistic Neural Network			76,0						
			SVM			75,0						
			CART			80,0						
			k Nearest neighbour			75,0						
			Naïve Bayes			89,0						
			Random forest			90,0						

Lists of abbreviations:

- LDA: Linear Discriminant Analysis
- MLP: Multilayer perceptron
- PNN: Probabilistic neural network
- SVM: Support Vector Machines
- CART: Classification and Regression Trees
- TP: True Positives values
- TN: True Negatives
- FP: False Positives
- FN: False Negatives

Evaluation criteria for models:

In the field of machine learning and in particular the problem of statistical classification, the performance of an algorithm is typically evaluated by a *confusion matrix*, also known as an *error matrix* (Chawla *et al.*, 2002) [19]. As the name stems, the confusion matrix makes it easy to see whether the system is confusing two classes, where "to confuse" is meant mislabeling one as another [9]. The table layout is organized (for a 2 class problem) as illustrated in Fig. 7: each row of the matrix represents the Actual class, while each column represents the Predicted class (or vice versa) (Chawla *et al.*, 2002) [19].

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Fig. 6: Confusion (error) Matrix (Chawla et al., 2002)

In a classification problem, each instance *I* is mapped to one element of the set {***p,n***}, where P are the positive instances and N are the negative instances for some condition. The predicted class is defined with the labels {***Y,N***} (Fawcett, 2006) [19].

Given a classifier and an instance, there are four possible outcomes (Fawcett, 2006; Chawla *et al.*, 2002; Chicco & Jurman, 2020);

- **True Positives (TP):** the number of positive examples correctly classified as positive;
- **False Negatives (FN):** the number of positive instances incorrectly classified as negative;
- **True Negatives (TN):** the number of negative examples correctly classified as negative;
- **False Positives (FP):** the number of negative examples **i**ncorrectly classified as positive.

Based on this matrix, many metrics have been defined, as shown in Fig. 8.

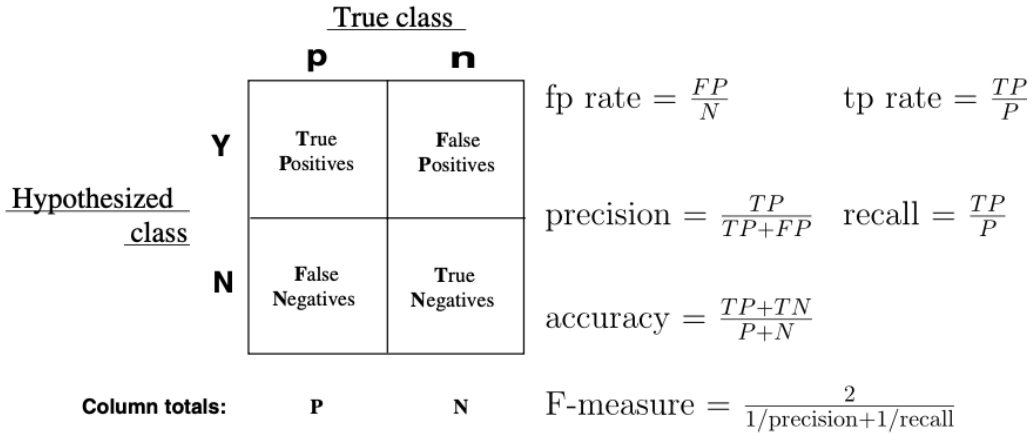


Fig. 7: confusion matrix and common performance metrics calculated from it (Fawcett, 2005)

In performing a machine learning binary classification, the goal is to maximize the number of true positives (TP) and true negatives (TN), and to minimize the number of false negatives (FN) and false positives (FP) (Chicco & Jurman, 2020). The numbers along the major diagonal represent correct predictions, while the number of the other diagonal represent the incorrect predictions, the errors (confusion) between the various classes (Fawcett, 2006; Powers, 2007).

"Since analyzing all the four categories of the confusion matrix separately would be time-consuming, statisticians introduced some useful statistical rates able to immediately describe the quality of a prediction" (Chicco & Jurman, 2020, p. 4).

Classifier comparison is traditionally based on *classification accuracy*, described as (Alduayi & Rajpoot, 2018; Lazzerini, 2021):

$$\text{Accuracy} = \frac{\text{correctly classified samples}}{\text{testing samples}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}$$

Alternatively, error rate = 1-accuracy.

In the context of balanced datasets and equal error costs, it is reasonable to use error rate as a performance metric (Chawla *et al.*, 2002). However, accuracy fails in providing a fair estimate of the classifier performance in the class-unbalanced datasets, as it assumes that the class prior probabilities are constant and relatively balanced and it assumes also equal error cost, a rarely case in the real world (Chicco & Jurman, 2020; Lazzerini 2021; Provost & Fawcett, 2001). "E.g., consider a domain where the classes appear in a 999:1 ratio. A simple rule "always classify as the maximum likelihood class" gives 99.9% accuracy (e.g., skews of 102 are common in fraud detection)" (Lazzerini, 2021, p.4).

Other common metrics that can be calculated from the confusion matrix are:

- **TP Rate (True Positive Rate):** defined by Alao & Adeyemo (2013) as "the proportion of cases which were classified as the actual class, indicating how much part of the class was correctly captured. It is equivalent to "Recall" and to "Sensitivity" [19].

  The TP Rate is estimated as (Alao & Adeyemo, 2013; Fawcett, 2006; Lazzerini, 2021; Provost & Fawcett, 2001):

  $$\text{TPR} = \frac{\text{positives correcy classified}}{\text{total positives}} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **Selectivity:** true positive rate (Ribes *et al.*, 2017)
- **TN Rate (True Negative Rate) or specificity** [19]**:** defined as (Fawcett, 2006; Chicco & Jurman, 2020):

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} = 1 - \mathrm{FPR}$$

- **FP Rate (False Positive Rate): t**he FP Rate (also called *false alarm rate*) of the classifier is estimated as (Alao & Adeyemo, 2013; Fawcett, 2006; Lazzerini, 2021; Provost & Fawcett, 2001):

$$\mathrm{FPR} = \frac{\text{negatives incorrecly classified}}{\text{total negatives}} = \frac{\mathrm{FP}}{\mathrm{N}} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} = 1 - \mathrm{TNR}$$

- **FN Rate (False Negative Rate):** also called "Miss rate" (Powers, 2007):

$$\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{P}} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}} = 1 - \mathrm{TPR}$$

Since the error rate is not a good metric for unbalanced datasets, the classification performance of algorithms in information retrieval is usually measured by precision and recall (Chawla *et al.*, 2002).

- **Precision:** denotes the proportion of Predicted Positive cases that are correctly Real Positives. It can analogously call be True Positive Accuracy (Powers, 2007). It is described as (Alduayi & Rajpoot, 2018)

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

- **Recall (or Sensitivity):** is the proportion of Real Positive cases that are correctly Predicted Positive (Powers, 2007). It is equivalent to TP Rate. It is described as (Alduayi & Rajpoot, 2018)

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

- **F1 Score/F-Measure:** defined as the harmonic mean of precision and recall (Alduayi & Rajpoot, 2018; Chicco & Jurman, 2020)

$$\text{F1 Score} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} = 2\frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

ROC graphs

A more general classifier comparison can be made with Receiver Operating Characteristic (ROC) analysis (Lazzerini, 2021; Provost & Fawcett, 2001). "ROC graphs are two-dimensional graphs in which tp rate is plotted on the Y axis and fp rate is plotted on the X axis. An ROC graph depicts relative tradeoffs between benefits (true positives) and costs (false positives)" (Fawcett, 2006); therefore this is a useful technique for visualizing and selecting classifier based on their

performance, especially in the presence of imbalanced datasets with unequal error costs (Lazzerini, 2021; Chawla *et al.*, 2002).

A discrete (binary) classifier (i.e., an instance of a confusion matrix) produces an (*fp rate, tp rate*) pair corresponding to a single point in ROC space (Fawcett, 2006; Lazzerini 2021).

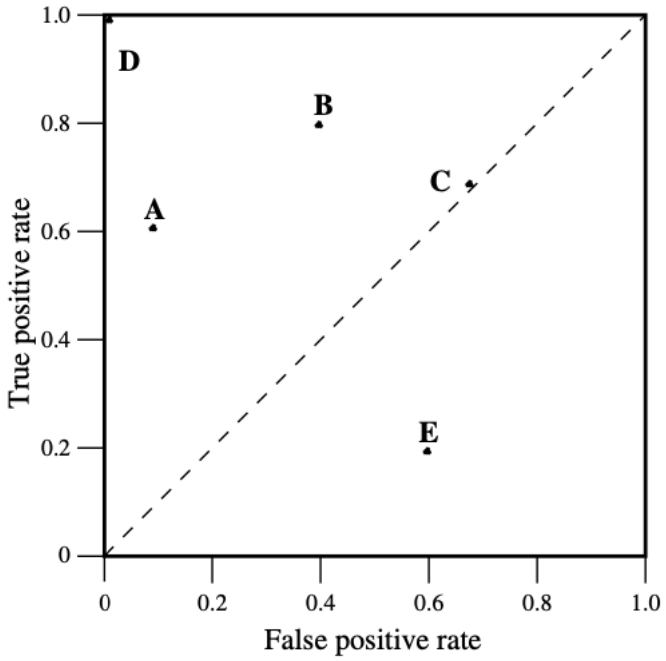


Fig. 8: a basic ROC graph showing five discrete classifiers (Fawcett, 2006).

An ROC graph allows a visual comparison of a set of classifiers (Lazzerini, 2021). One point in ROC space is better than another if it is to the northwest (higher TPR, lower FPR, or both) of the first. The diagonal line y = x represents the strategy of randomly guessing a class (any classifier that appears in the lower right triangle performs worse than random guessing, and for this reason this triangle is usually empty in ROC graphs) (Fawcett, 2006; Lazzerini, 2021).

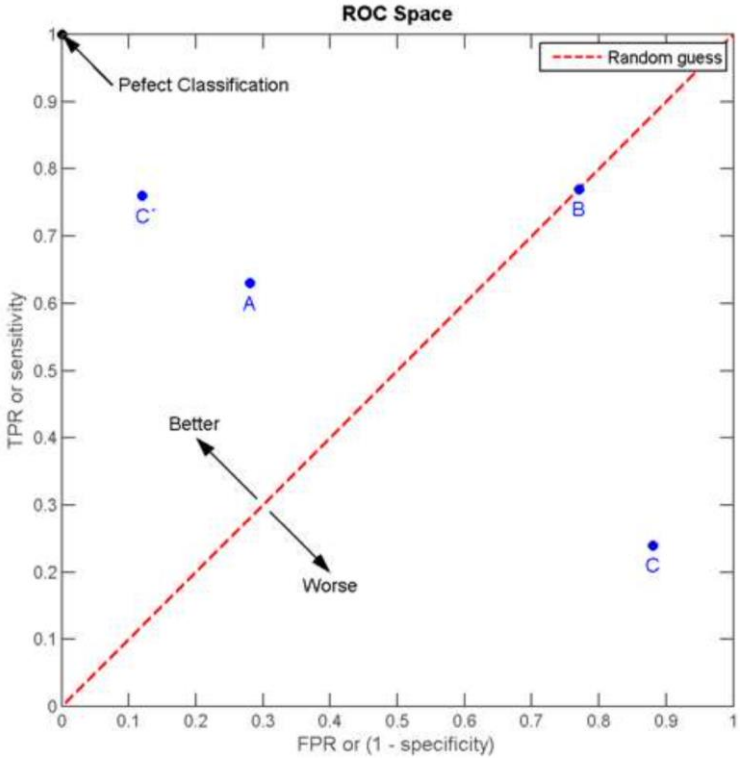


Fig. 9: visual comparison of a set of classifiers (Lazzerini, 2021).

Curves in the ROC space

In continuous classifiers, such as Naïve Bayes or neural networks, forecasts are expressed in terms of the probability that an instance is a member of a positive or a negative class (Mason & Graham, 2002; Fawcett, 2006). The probability is related to a specific choice of the threshold: if the classifier output is above the threshold, the classifier produces *Y*, else *N* (Lazzerini, 2021; Fawcett, 2006). The probability varies across a range of thresholds, for each threshold is produced a different point in the ROC space (each threshold defines a classifier) (Mason & Graham, 2002; Fawcett, 2006; Provost & Fawcett, 2001). These points collectively define the so-called *ROC curve* and it illustrates the error tradeoffs available with a given model (Mason & Graham, 2002; Lazzerini, 2021; Provost & Fawcett, 2001).

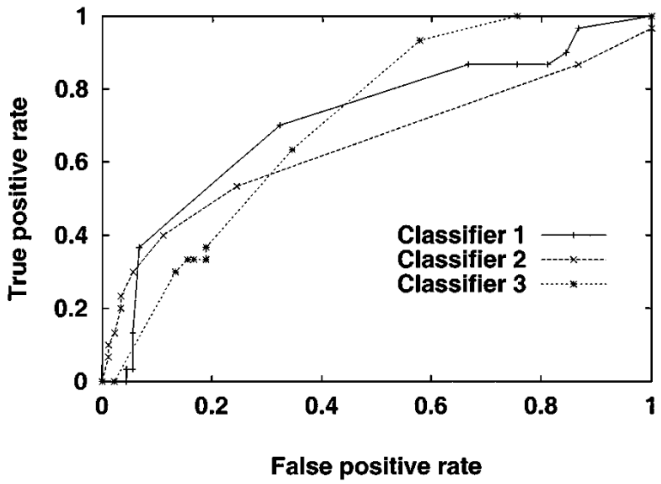


Fig. 10: ROC graphs of three classifiers (Provost & Fawcett, 2001).

The ROC graph is a valuable visualization technique, but it can help to choose the best global classifier only when one classifier dominates another over the entire performance space (Provost & Fawcett, 2001; Lazzerini, 2021).

Area Under the Curve (AUC)

If no classifier dominates all others over the entire ROC space, some researchers adopted the method of choosing the classifier that maximizes the Area Under the (ROC) Curve (AUC). It represents the average performance over the entire curve, in the case where there is no interest in a specific trade-off between TPR and FPR. It is particularly used in situations where either the target cost distribution or class distribution is completely unknown (Provost & Fawcett, 2001; Lazzerini, 2021).

4.2 Model Building

The modelling process consists in selecting models on the basis of the results of the literature experiments described in the previous chapter, with the aim to identify the best classifier for the analysed problem.

We therefore selected four classifiers, based on the performance evidence shown in Table 3 and our choice to analyze a couple of simpler (Naïve Bayes[16] and Logistic Regression) and a couple more complex methods (Decision Tree[17] and Random Forest[18]). Although the Logistic Regression was not

[16] Majka M (2019). *naivebayes: High Performance Implementation of the Naive Bayes Algorithm in R*. R package version 0.9.7, https://CRAN.R-project.org/package=naivebayes.

[17] Milborrow, S. (2020). Plotting rpart trees with the rpart.plot package. http://www.milbo.org/rpart-plot/prp.pdf

[18] Liaw A. and Wiener M. (2002). Classification and Regression by randomForest. R News, 3(3), 18-22. https://CRAN.R-project.org/doc/Rnews/

recommended by the literature, we were interested to see how it behaved compared to other mainstream methods.

We trained each classifier on the featured set and the classifier with the best classification results is used for prediction.

The classification algorithms taken into consideration are:

- Logistic Regression,
- Naïve Bayes,
- Decision Tree,
- Random Forest.

Table 4: strenghts and weaknesses of the machine learning algorithms considered (source: Lantz (2019))

Algorithm	Strengths	Weaknesses	Main applications
Logistic Regression	• By far the most common approach for modeling numeric data • Can be adapted to model almost any modeling task • Provides estimates of both the size and strength of the relationships among features and the outcome • They are computationally simple algorithms	• Makes strong assumptions about the data • The model's form must be specified by the user in advance • Does not handle missing data • Only works with numeric features, so categorical data requires additional preparation (reason why we used regularizations) • Requires some knowledge of statistics to understand the model	• Examining how populations and individuals vary by their measured characteristics • Quantifying the causal relationship between an event and its response • Identifying patterns that can be used to forecast future behavior given known criteria • Statistical hypothesis testing
Naïve Bayes	• Simple, fast and very effective (it assumes that all the features in the dataset are equally important and independent) • Does well with noisy and missing data • Requires relatively few examples for training, but also works well with very large numbers of examples • Easy to obtain the estimated probability for a prediction • They are computationally simple algorithms	• Relies on an often-faulty assumption of equally important and independent features • Not ideal for datasets with many numeric features • Estimated probabilities are less reliable than the predicted classes	• Text classification (for example: identify spam by monitoring email messages)

Decision Tree	• The algorithm output the resulting structure in a human-readable format • An all-purpose classifier that can be applied for modeling many types of problems and almost any type of data • More efficient that other complex models	• Tendency to overfit data • Are often biased towards splits on features having a large number of levels	• Credit scoring modeling • Marketing studies of customer behaviour • Diagnosis of medical conditions
Random Forest	• They are more accurate than decision tree • An all-purpose model that performs well on most problems • Can handle noisy or missing data, as well as categorical or continuous features • Select only the most important features • Can be used on data with an extreme large number of features or examples	• Unlike decision tree, the model is not easily interpretable • They are computationally expensive	• Extremely large datasets

In all models considered, the dataset was randomly divided first into an 80 % training set and 20 % test set, and then into an 70 % training set and 30 % test set. All models were trained using ten-fold cross validation on the training set. The trained models from each algorithm were then used to predict and test on the 20/30 % test set.

Finally, we used the ROSE method to address the issue of class imbalance, and we implemented it though the ROSE package in R. “The ROSE package provides functions to deal with binary classification problems in the presence of imbalanced classes” (Lunardon *et al.*, 2014).

We removed the variable "TitoloStudio", “Potenziale”, “Clima”, “ClasseNumeroFigli” and “Leadership” from the models because most of the cases in which the value was absent concerned the employees who resigned, and this coincidence affected the predictions. This happens because, as regards the variables that represent a description of personal characteristics, ("Titolostudio" and "ClasseNumeroFigli") the company policy provides that, when the employee leaves the organization, the relative information is deleted, while for “Potenziale”, “Clima” and “Leadership” it most of the employees who left the organization have not had this type of assessment. The variable “Ruolo” has been replaced with the job decription it represents (“Job_description”).

We therefore used the remaining 14 variables as inputs for the predictive models:

1. Genere;
2. ClasseStruttura;
3. Posizione;
4. Orario;
5. Grado;

6. Contratto;
7. Talento;
8. Gratifica;
9. Retention;
10. ClasseEta;
11. ClasseAnniServizio;
12. Job_description;
13. LogRetribuzione;
14. LogPremio.

4.2.1 Logistic Regression

Regularization pars down independent variables that have low predictive value[19].

We used three different regularisations, which serve to prevent the algorithm, in the presence of a large number of regressors, had poor performance:

- Lasso: penalty based on the L1-norm;
- Ridge: penalty based on the L2-norm;
- Elastic: a combination of the above.

We set the optimal penalty (λ) by 10-fold cross-validation (one standard deviation criterion) and maximize AUC.

We apply weights to rebalance Leave:Stay to 50:50 because logistic regression performs best when data are balanced.

80:20 split

LASSO

[19] https://www.datacamp.com/community/tutorials/tutorial-ridge-lasso-elastic-net

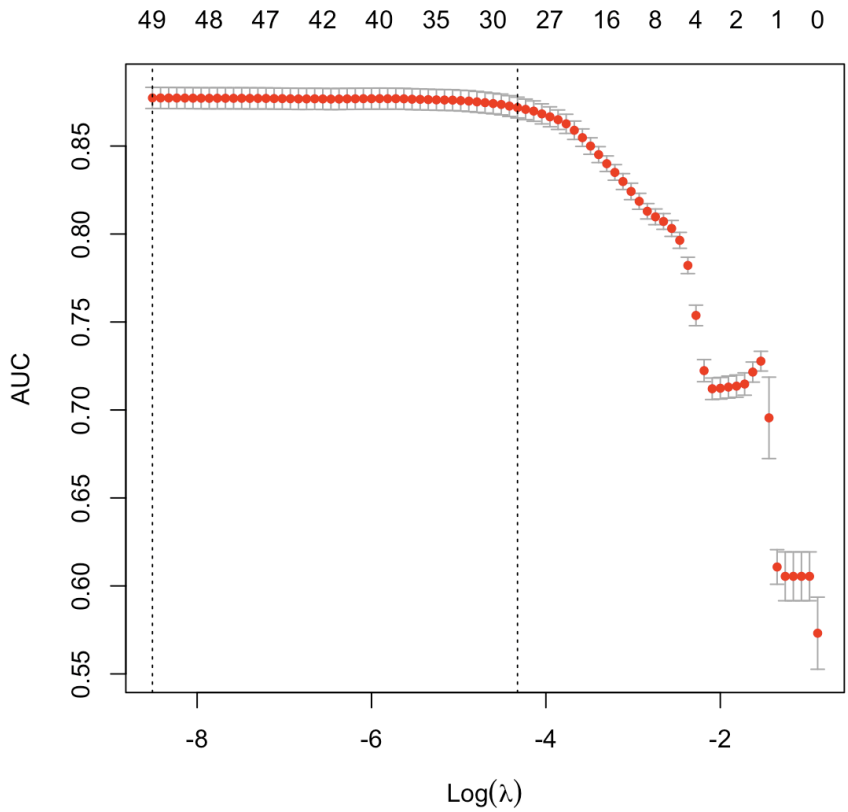


Fig. 11: AUC maximization for lambda values

```
Call:  cv.glmnet(x = X, y = Y, type.measure = "auc", nfolds = 10, intercept = FALSE,      fa
mily = "binomial", alpha = 1)

Measure: AUC

      Lambda Measure       SE Nonzero
min 0.000201  0.8773 0.006005      49
1se 0.013224  0.8719 0.005906      28
```

Fig. 12: results for the AUC maximization for lambda values

As can be seen from the figures (above), the area below the curve is maximized for low levels of lambda.

The figure below shows the estimated regression parameters. The results of the dots show that the values are so small that they are considered irrelevant. Sign shows whether the effect on the probability of resign is positive or negative.

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                      1
(Intercept)                                    .
GenereFemmina                                  .
GenereMaschio                                  2.683730e-01
ClasseStrutturaRete                           -3.302019e-01
ClasseStrutturaSocietà                         .
PosizioneResponsabile                         -3.081598e-02
OrarioPart Time                                .
Grado2                                         .
Grado3                                         .
Grado4                                        -2.623274e-01
Grado5                                         2.303747e-01
ContrattoCredito                              -1.047123e-01
ContrattoDirigenti                             3.691318e-15
ContrattoFunzionari                            .
TalentoSì                                      .
GratificaSì                                    .
RetentionSì                                    1.486838e+00
ClasseEta(25, 30]                              9.109537e-03
ClasseEta(30, 40]                              .
ClasseEta(40, 50]                             -8.188238e-01
ClasseEta(50, +Inf]                           -1.252566e+00
ClasseAnniServizio(3,10]                       1.283059e-01
ClasseAnniServizio(10,20]                     -8.209680e-01
ClasseAnniServizio(20,30]                     -1.133373e+00
ClasseAnniServizio(30,Inf]                    -1.263490e+00
Job_descriptionAREA CREDITI                   -7.396457e-01
Job_descriptionAREA MERCATI                    5.101770e-01
Job_descriptionAREA TECNICA E LOGISTICA        .
Job_descriptionAUDIT E COMPLIANCE              .
Job_descriptionBACK OFFICE E AMMINISTRAZIONE   .
Job_descriptionCONSULENZA FINANZIARIA          8.317559e-01
Job_descriptionCORPORATE BANKING               .
Job_descriptionCUSTOMER ASSISTANCE            -3.517857e-01
```

Fig. 13: LASSO regression estimated parameters

RIDGE

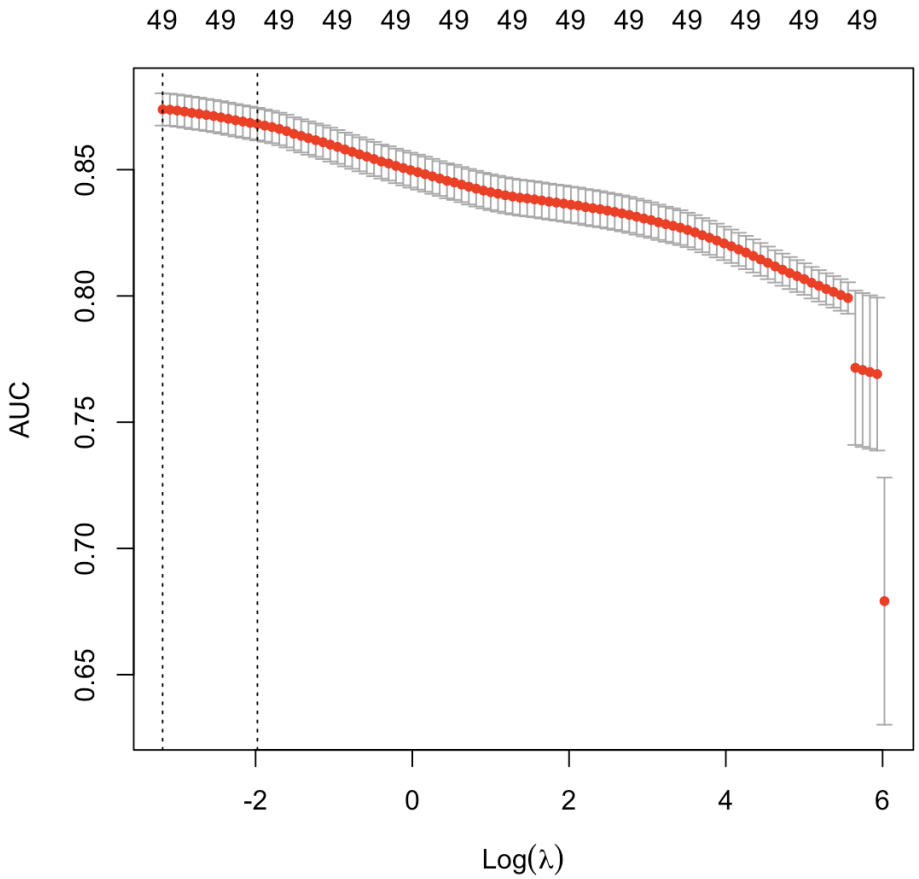


Fig. 14: AUC maximization for lambda values

```
Call:  cv.glmnet(x = X, y = Y, type.measure = "auc", nfolds = 10, intercept = FALSE,      fa
mily = "binomial", alpha = 0)

Measure: AUC

     Lambda Measure       SE Nonzero
min 0.04133  0.8739 0.006373      49
1se 0.13853  0.8681 0.006431      49
```

Fig. 15: results for the AUC maximization for lambda values

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                    1
(Intercept)                                 .
GenereFemmina                              -0.110949329
GenereMaschio                               0.140035133
ClasseStrutturaRete                        -0.141148192
ClasseStrutturaSocietà                      0.436511183
PosizioneResponsabile                      -0.333926476
OrarioPart Time                             0.050903163
Grado2                                     -0.089666804
Grado3                                     -0.061247485
Grado4                                     -0.382039387
Grado5                                      0.156963832
ContrattoCredito                           -0.105852803
ContrattoDirigenti                          0.156997588
ContrattoFunzionari                         0.039690191
TalentoSì                                  -0.003566983
GratificaSì                                -0.120872535
RetentionSì                                 1.068224552
ClasseEta(25, 30]                           0.307255325
ClasseEta(30, 40]                           0.173283754
ClasseEta(40, 50]                          -0.416675079
ClasseEta(50, +Inf]                        -0.688980498
ClasseAnniServizio(3,10]                    0.297924754
ClasseAnniServizio(10,20]                  -0.460692904
ClasseAnniServizio(20,30]                  -0.663585342
ClasseAnniServizio(30,Inf]                 -0.814467115
Job_descriptionAREA CREDITI                -0.865644777
Job_descriptionAREA MERCATI                 0.879051056
Job_descriptionAREA TECNICA E LOGISTICA    -0.146180949
Job_descriptionAUDIT E COMPLIANCE          -0.043482687
Job_descriptionBACK OFFICE E AMMINISTRAZIONE  0.014033365
Job_descriptionCONSULENZA FINANZIARIA       0.434988933
Job_descriptionCORPORATE BANKING           -0.155098099
Job_descriptionCUSTOMER ASSISTANCE         -0.345668857
Job_descriptionDIREZIONE                    1.220925389
Job_descriptionFINANZA                     -0.147420776
Job_descriptionGOVERNANCE E ORGANIZZAZIONE  -0.128664995
Job_descriptionICT                          0.060445783
Job_descriptionINSURANCE                   -1.598204495
Job_descriptionLEGAL                        0.555479274
Job_descriptionPEOPLE MANAGEMENT           -0.319563026
Job_descriptionRETAIL BANKING              -0.194474998
Job_descriptionRISK E AL MANAGEMENT         0.470694318
Job_descriptionSERVIZI CONDIVISI           -0.826368810
Job_descriptionSERVIZI DI CASSA            -0.149164478
Job_descriptionSICUREZZA E PREVENZIONE     -1.014838405
Job_descriptionSTAFF DI DIREZIONE          -1.521266101
Job_descriptionSVILUPPO ASSET               0.647139534
Job_descriptionSVILUPPO CLIENTI             0.360858635
LogRetribuzione                             0.287393963
LogPremio                                  -0.311987804
```

Fig. 16: RIDGE regression estimated parameters

MIX

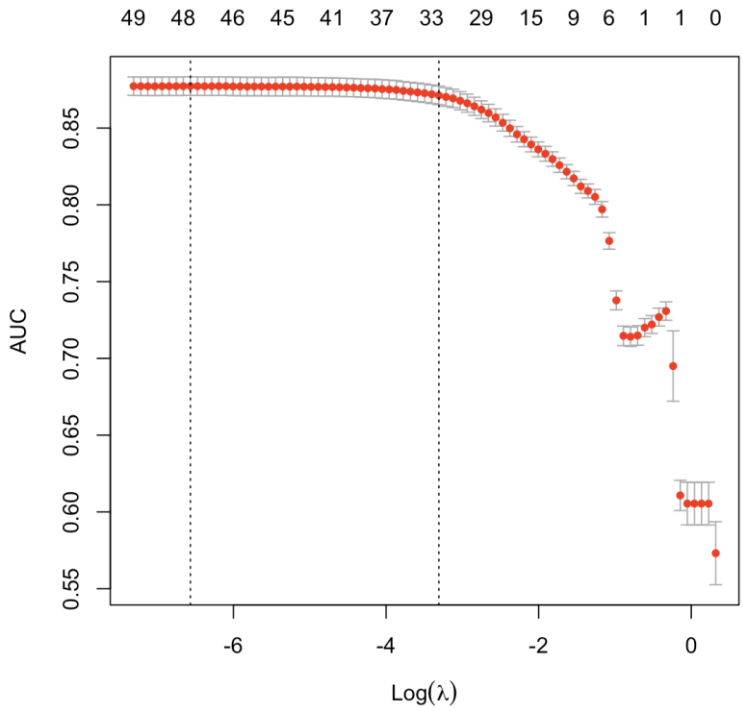


Fig. 17: AUC maximization for lambda values

```
Call:  cv.glmnet(x = X, y = Y, type.measure = "auc", nfolds = 10, intercept = FALSE,      fa
mily = "binomial", alpha = 0.3)

Measure: AUC

     Lambda Measure       SE Nonzero
min 0.00141  0.8773 0.006044      48
1se 0.03659  0.8713 0.006061      32
```

Fig. 18: results for the AUC maximization for lambda values

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                    1
(Intercept)                                         .
GenereFemmina                                -0.05758973
GenereMaschio                                 0.19030495
ClasseStrutturaRete                          -0.23739383
ClasseStrutturaSocietà                        .
PosizioneResponsabile                        -0.15962988
OrarioPart Time                               .
Grado2                                        .
Grado3                                        .
Grado4                                       -0.30451178
Grado5                                        0.13170660
ContrattoCredito                             -0.10562284
ContrattoDirigenti                            0.13200123
ContrattoFunzionari                           .
TalentoSì                                     .
GratificaSì                                   .
RetentionSì                                   1.31958265
ClasseEta(25, 30]                             0.16663600
ClasseEta(30, 40]                             0.06301350
ClasseEta(40, 50]                            -0.62403171
ClasseEta(50, +Inf]                          -0.99548001
ClasseAnniServizio(3,10]                      0.21516215
ClasseAnniServizio(10,20]                    -0.65606552
ClasseAnniServizio(20,30]                    -0.94017949
ClasseAnniServizio(30,Inf]                   -1.07439960
Job_descriptionAREA CREDITI                  -0.77259801
Job_descriptionAREA MERCATI                   0.65340258
Job_descriptionAREA TECNICA E LOGISTICA       .
Job_descriptionAUDIT E COMPLIANCE             .
Job_descriptionBACK OFFICE E AMMINISTRAZIONE  .
Job_descriptionCONSULENZA FINANZIARIA         0.67310542
Job_descriptionCORPORATE BANKING              .
Job_descriptionCUSTOMER ASSISTANCE           -0.32628116
Job_descriptionDIREZIONE                      1.85020707
Job_descriptionFINANZA                        .
Job_descriptionGOVERNANCE E ORGANIZZAZIONE    .
Job_descriptionICT                            .
Job_descriptionINSURANCE                     -0.13812438
Job_descriptionLEGAL                          0.53978581
Job_descriptionPEOPLE MANAGEMENT              .
Job_descriptionRETAIL BANKING                -0.04352146
Job_descriptionRISK E AL MANAGEMENT           0.45256828
Job_descriptionSERVIZI CONDIVISI             -0.96476536
Job_descriptionSERVIZI DI CASSA              -0.14397604
Job_descriptionSICUREZZA E PREVENZIONE        .
Job_descriptionSTAFF DI DIREZIONE             .
Job_descriptionSVILUPPO ASSET                 0.68429317
Job_descriptionSVILUPPO CLIENTI               0.29518622
LogRetribuzione                               0.41898010
LogPremio                                    -0.45966675
```

Fig. 19: MIX regression estimated parameters

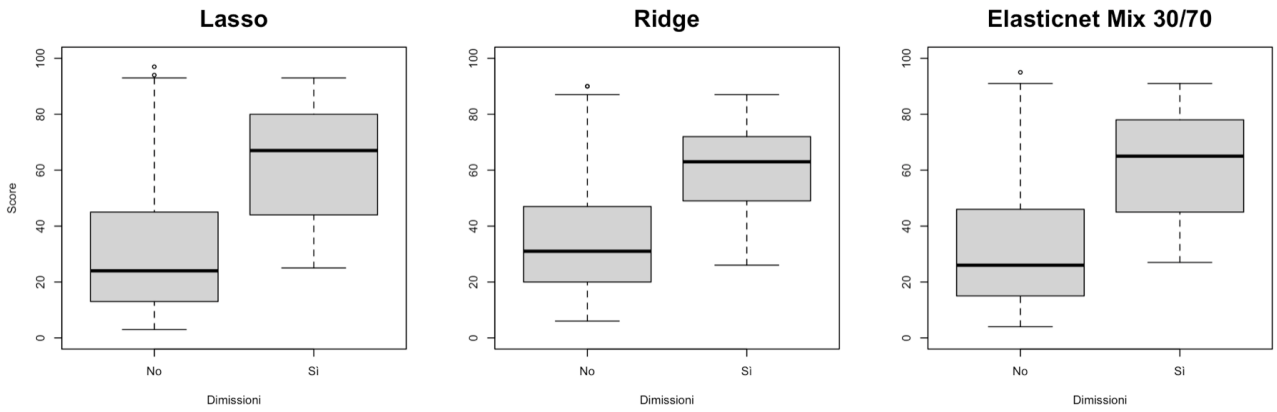


Fig. 20: distribution of the Score on the test set of the three models

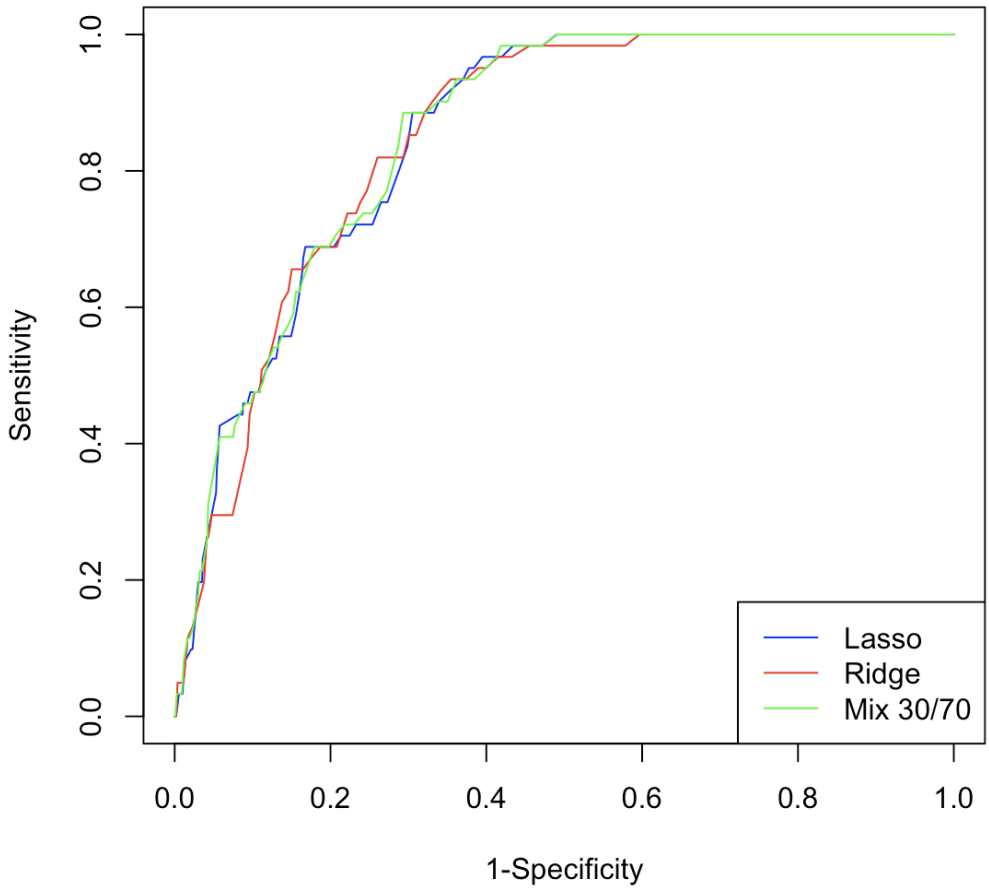


Fig. 21: ROC curve of the three models

From the graph above we can see that the ROC curves of ELASTCNET and LASSO are quite similar, that of RIDGE is slightly different.

```
Lasso
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No  Sì
        No 868  19
        Sì 223  42

               Accuracy : 0.7899
                 95% CI : (0.7652, 0.8131)
    No Information Rate : 0.947
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1877

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.68852
            Specificity : 0.79560
         Pos Pred Value : 0.15849
         Neg Pred Value : 0.97858
             Prevalence : 0.05295
         Detection Rate : 0.03646
   Detection Prevalence : 0.23003
      Balanced Accuracy : 0.74206

       'Positive' Class : Sì
```

Fig. 22: LASSO Confusion Matrix and Statistics

```
Ridge
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No  Sì
        No 837  16
        Sì 254  45

               Accuracy : 0.7656
                 95% CI : (0.7401, 0.7898)
    No Information Rate : 0.947
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1777

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.73770
            Specificity : 0.76719
         Pos Pred Value : 0.15050
         Neg Pred Value : 0.98124
             Prevalence : 0.05295
         Detection Rate : 0.03906
   Detection Prevalence : 0.25955
      Balanced Accuracy : 0.75245

       'Positive' Class : Sì
```

Fig. 23: RIDGE Confusion Matrix and Statistics

```
Elasticnet Mix 30/70
```

```
Confusion Matrix and Statistics

          Reference
Prediction  No  Sì
        No 866  18
        Sì 225  43

               Accuracy : 0.7891
                 95% CI : (0.7643, 0.8123)
    No Information Rate : 0.947
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1917

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.70492
            Specificity : 0.79377
         Pos Pred Value : 0.16045
         Neg Pred Value : 0.97964
             Prevalence : 0.05295
         Detection Rate : 0.03733
   Detection Prevalence : 0.23264
      Balanced Accuracy : 0.74934

       'Positive' Class : Sì
```

Fig. 24: Elasticnet Confusion Matrix and Statistics

70:30 split

LASSO

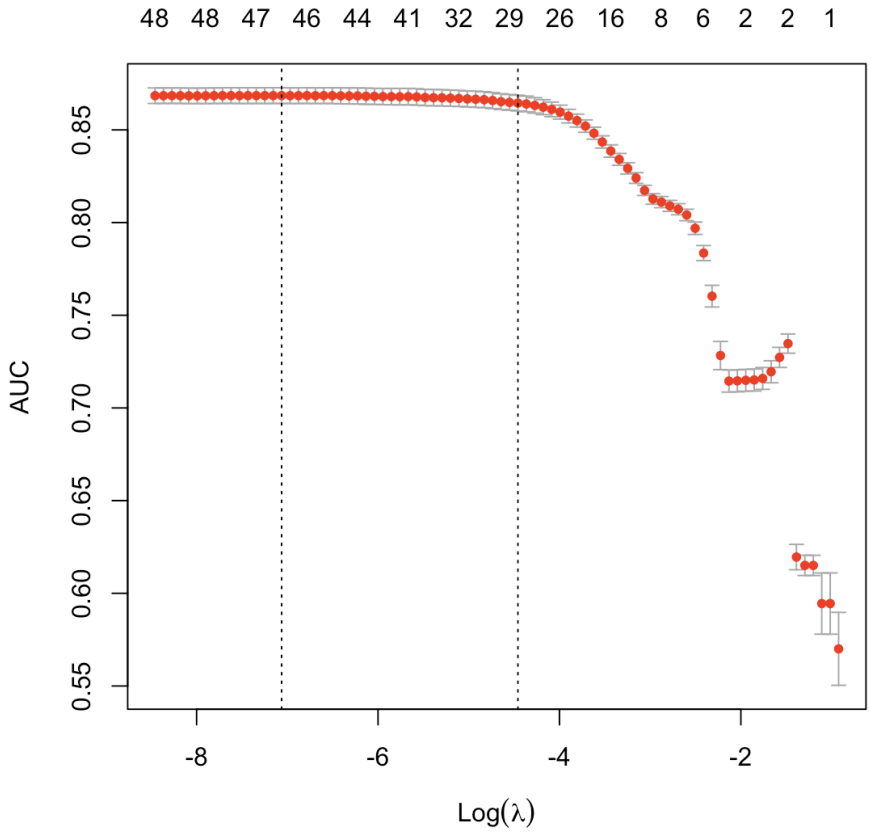


Fig. 25: AUC maximization for lambda values

```
Call:  cv.glmnet(x = X, y = Y, type.measure = "auc", nfolds = 10, intercept = FALSE,      fa
mily = "binomial", alpha = 1)

Measure: AUC

      Lambda Measure       SE Nonzero
min 0.000857  0.8685 0.004231      47
1se 0.011589  0.8645 0.004241      29
```

Fig. 26: results for the AUC maximization for lambda values

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                  1
(Intercept)                                .
GenereFemmina                              .
GenereMaschio                               0.2707445557
ClasseStrutturaRete                        -0.3567605968
ClasseStrutturaSocietà                     .
PosizioneResponsabile                      -0.0080667431
OrarioPart Time                            .
Grado2                                     .
Grado3                                     .
Grado4                                     -0.3960145567
Grado5                                      0.3362205217
ContrattoCredito                           -0.0233385988
ContrattoDirigenti                          0.0003932901
ContrattoFunzionari                        .
TalentoSì                                  .
GratificaSì                                .
RetentionSì                                 1.5746928165
ClasseEta(25, 30]                          .
ClasseEta(30, 40]                          .
ClasseEta(40, 50]                          -0.8101473126
ClasseEta(50, +Inf]                        -1.0789426890
ClasseAnniServizio(3,10]                    0.2219857914
ClasseAnniServizio(10,20]                  -0.6424938395
ClasseAnniServizio(20,30]                  -1.0950312775
ClasseAnniServizio(30,Inf]                 -1.3733226865
Job_descriptionAREA CREDITI                -1.0467777775
Job_descriptionAREA MERCATI                 0.4937764219
Job_descriptionAREA TECNICA E LOGISTICA    .
Job_descriptionAUDIT E COMPLIANCE          .
Job_descriptionBACK OFFICE E AMMINISTRAZIONE .
Job_descriptionCONSULENZA FINANZIARIA       0.7871735565
Job_descriptionCORPORATE BANKING           .
Job_descriptionCUSTOMER ASSISTANCE         -0.3796731840
Job_descriptionDIREZIONE                    2.3096539455
Job_descriptionFINANZA                     .
Job_descriptionGOVERNANCE E ORGANIZZAZIONE  .
Job_descriptionICT                         .
Job_descriptionINSURANCE                   -0.0441326097
Job_descriptionLEGAL                        0.3803148700
Job_descriptionPEOPLE MANAGEMENT           -0.0548663194
Job_descriptionRETAIL BANKING              .
Job_descriptionRISK E AL MANAGEMENT         0.3094380408
Job_descriptionSERVIZI CONDIVISI           -1.1916622505
Job_descriptionSERVIZI DI CASSA            -0.1449256969
Job_descriptionSICUREZZA E PREVENZIONE     .
Job_descriptionSTAFF DI DIREZIONE          .
Job_descriptionSVILUPPO ASSET               0.8426180253
Job_descriptionSVILUPPO CLIENTI             0.3424111476
LogRetribuzione                             0.4852723030
LogPremio                                  -0.5387876894
```

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                      1
(Intercept)                                           .
GenereFemmina                                         .
GenereMaschio                                         0.2707445557
ClasseStrutturaRete                                  -0.3567605968
ClasseStrutturaSocietà                                .
PosizioneResponsabile                                -0.0080667431
OrarioPart Time                                       .
Grado2                                                .
Grado3                                                .
Grado4                                               -0.3960145567
Grado5                                                0.3362205217
ContrattoCredito                                     -0.0233385988
ContrattoDirigenti                                    0.0003932901
ContrattoFunzionari                                   .
TalentoSì                                             .
GratificaSì                                           .
RetentionSì                                           1.5746928165
ClasseEta(25, 30]                                     .
ClasseEta(30, 40]                                     .
ClasseEta(40, 50]                                    -0.8101473126
ClasseEta(50, +Inf]                                  -1.0789426890
ClasseAnniServizio(3,10]                              0.2219857914
ClasseAnniServizio(10,20]                            -0.6424938395
ClasseAnniServizio(20,30]                            -1.0950312775
ClasseAnniServizio(30,Inf]                           -1.3733226865
Job_descriptionAREA CREDITI                          -1.0467777775
Job_descriptionAREA MERCATI                           0.4937764219
Job_descriptionAREA TECNICA E LOGISTICA               .
Job_descriptionAUDIT E COMPLIANCE                     .
Job_descriptionBACK OFFICE E AMMINISTRAZIONE          .
Job_descriptionCONSULENZA FINANZIARIA                 0.7871735565
Job_descriptionCORPORATE BANKING                      .
Job_descriptionCUSTOMER ASSISTANCE                   -0.3796731840
Job_descriptionDIREZIONE                              2.3096539455
Job_descriptionFINANZA                                .
Job_descriptionGOVERNANCE E ORGANIZZAZIONE            .
Job_descriptionICT                                    .
Job_descriptionINSURANCE                             -0.0441326097
Job_descriptionLEGAL                                  0.3803148700
Job_descriptionPEOPLE MANAGEMENT                     -0.0548663194
Job_descriptionRETAIL BANKING                         .
Job_descriptionRISK E AL MANAGEMENT                   0.3094380408
Job_descriptionSERVIZI CONDIVISI                     -1.1916622505
Job_descriptionSERVIZI DI CASSA                      -0.1449256969
Job_descriptionSICUREZZA E PREVENZIONE                .
Job_descriptionSTAFF DI DIREZIONE                     .
Job_descriptionSVILUPPO ASSET                         0.8426180253
Job_descriptionSVILUPPO CLIENTI                       0.3424111476
LogRetribuzione                                       0.4852723030
LogPremio                                            -0.5387876894
```

Fig. 27: LASSO regression estimated parameters

RIDGE

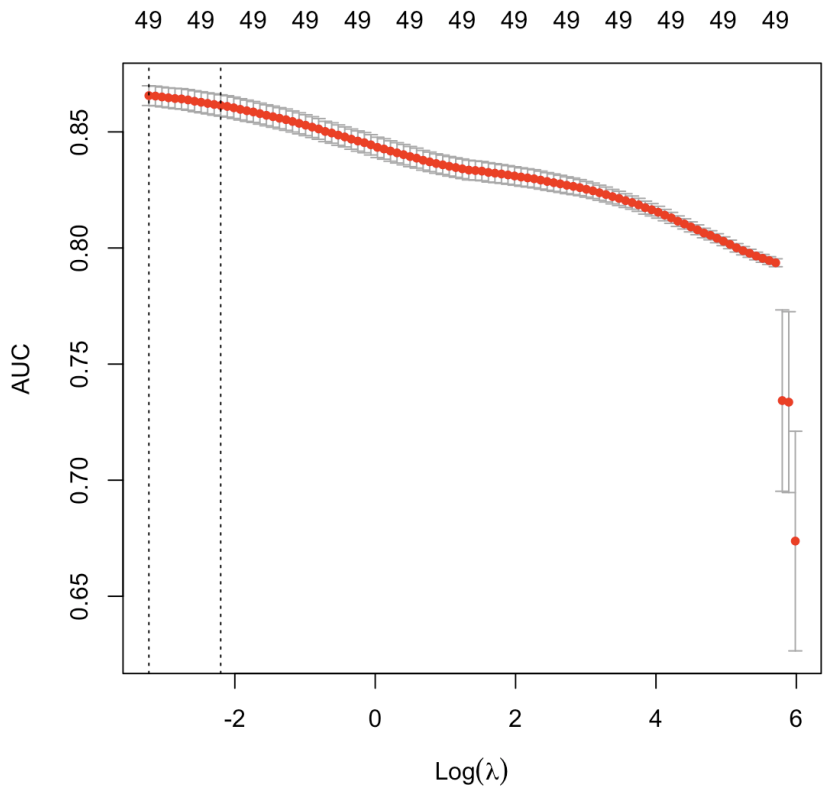


Fig. 28: AUC maximization for lambda values

```
Call:  cv.glmnet(x = X, y = Y, type.measure = "auc", nfolds = 10, intercept = FALSE,      fa
mily = "binomial", alpha = 0)

Measure: AUC

     Lambda Measure       SE Nonzero
min 0.03976  0.8656 0.004249      49
1se 0.11062  0.8614 0.004590      49
```

Fig. 29: results for the AUC maximization for lambda values

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                      1
(Intercept)                                           .
GenereFemmina                                -0.110241919
GenereMaschio                                 0.142954749
ClasseStrutturaRete                          -0.132691205
ClasseStrutturaSocietà                        0.470919210
PosizioneResponsabile                        -0.323549686
OrarioPart Time                               0.062075086
Grado2                                       -0.109068306
Grado3                                       -0.044843514
Grado4                                       -0.463031031
Grado5                                        0.179206663
ContrattoCredito                             -0.104619798
ContrattoDirigenti                            0.179293340
ContrattoFunzionari                           0.009793757
TalentoSì                                    -0.041735334
GratificaSì                                  -0.188062636
RetentionSì                                   1.134976125
ClasseEta(25, 30]                             0.292356329
ClasseEta(30, 40]                             0.193350372
ClasseEta(40, 50]                            -0.458445193
ClasseEta(50, +Inf]                          -0.676005029
ClasseAnniServizio(3,10]                      0.316660153
ClasseAnniServizio(10,20]                    -0.425514337
ClasseAnniServizio(20,30]                    -0.708826324
ClasseAnniServizio(30,Inf]                   -0.909757700
Job_descriptionAREA CREDITI                  -1.018195923
Job_descriptionAREA MERCATI                   0.945537315
Job_descriptionAREA TECNICA E LOGISTICA       0.005038068
Job_descriptionAUDIT E COMPLIANCE            -0.048257202
Job_descriptionBACK OFFICE E AMMINISTRAZIONE -0.030877176
Job_descriptionCONSULENZA FINANZIARIA         0.440495008
Job_descriptionCORPORATE BANKING             -0.259397496
Job_descriptionCUSTOMER ASSISTANCE           -0.397669444
Job_descriptionDIREZIONE                      1.451481714
Job_descriptionFINANZA                       -0.084556157
Job_descriptionGOVERNANCE E ORGANIZZAZIONE    0.006395457
Job_descriptionICT                            0.073424174
Job_descriptionINSURANCE                     -1.830452243
Job_descriptionLEGAL                          0.501312818
Job_descriptionPEOPLE MANAGEMENT             -0.690339364
Job_descriptionRETAIL BANKING                -0.196077207
Job_descriptionRISK E AL MANAGEMENT           0.433731145
Job_descriptionSERVIZI CONDIVISI             -0.884059775
Job_descriptionSERVIZI DI CASSA              -0.145211393
Job_descriptionSICUREZZA E PREVENZIONE       -1.128017934
Job_descriptionSTAFF DI DIREZIONE            -1.012409374
Job_descriptionSVILUPPO ASSET                 0.709007917
Job_descriptionSVILUPPO CLIENTI               0.364723240
LogRetribuzione                               0.306323919
LogPremio                                    -0.338550325
```

Fig. 30: RIDGE regression estimated parameters

MIX

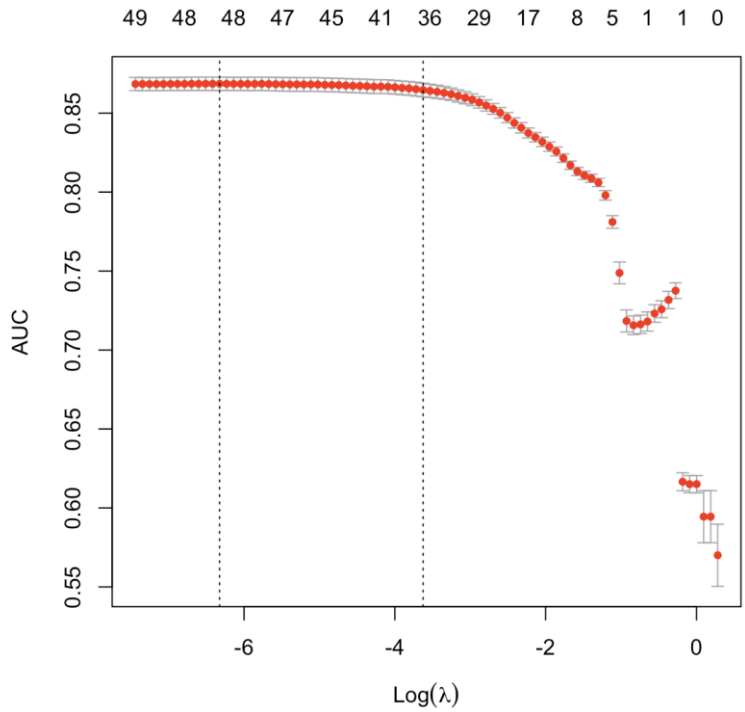


Fig. 31: AUC maximization for lambda values

```
Call:  cv.glmnet(x = X, y = Y, type.measure = "auc", nfolds = 10, intercept = FALSE,      fa
mily = "binomial", alpha = 0.3)

Measure: AUC

      Lambda Measure       SE Nonzero
min 0.001793  0.8686 0.004224      48
1se 0.026626  0.8646 0.004334      36
```

Fig. 32: results for the AUC maximization for lambda values

```
50 x 1 sparse Matrix of class "dgCMatrix"
                                                    1
(Intercept)                                         .
GenereFemmina                                -0.05451638
GenereMaschio                                 0.21543825
ClasseStrutturaRete                          -0.27585385
ClasseStrutturaSocietà                        .
PosizioneResponsabile                        -0.16484295
OrarioPart Time                               .
Grado2                                       -0.01101282
Grado3                                        .
Grado4                                       -0.43490200
Grado5                                        0.19851678
ContrattoCredito                             -0.07326911
ContrattoDirigenti                            0.19907318
ContrattoFunzionari                           .
TalentoSì                                     .
GratificaSì                                  -0.03620926
RetentionSì                                   1.49262355
ClasseEta(25, 30]                             0.12369705
ClasseEta(30, 40]                             0.08935965
ClasseEta(40, 50]                            -0.70214139
ClasseEta(50, +Inf]                          -0.96858608
ClasseAnniServizio(3,10]                      0.25533323
ClasseAnniServizio(10,20]                    -0.59229718
ClasseAnniServizio(20,30]                    -0.98818069
ClasseAnniServizio(30,Inf]                   -1.25650568
Job_descriptionAREA CREDITI                  -1.14454377
Job_descriptionAREA MERCATI                   0.75567950
Job_descriptionAREA TECNICA E LOGISTICA       .
Job_descriptionAUDIT E COMPLIANCE             .
Job_descriptionBACK OFFICE E AMMINISTRAZIONE  .
Job_descriptionCONSULENZA FINANZIARIA         0.66683600
Job_descriptionCORPORATE BANKING             -0.04952195
Job_descriptionCUSTOMER ASSISTANCE           -0.42854742
Job_descriptionDIREZIONE                      2.16772358
Job_descriptionFINANZA                        .
Job_descriptionGOVERNANCE E ORGANIZZAZIONE    .
Job_descriptionICT                            .
Job_descriptionINSURANCE                     -0.63609669
Job_descriptionLEGAL                          0.45609938
Job_descriptionPEOPLE MANAGEMENT             -0.34054266
Job_descriptionRETAIL BANKING                -0.07635322
Job_descriptionRISK E AL MANAGEMENT           0.37721910
Job_descriptionSERVIZI CONDIVISI             -1.15419106
Job_descriptionSERVIZI DI CASSA              -0.17376942
Job_descriptionSICUREZZA E PREVENZIONE        .
Job_descriptionSTAFF DI DIREZIONE             .
Job_descriptionSVILUPPO ASSET                 0.81718495
Job_descriptionSVILUPPO CLIENTI               0.35059076
LogRetribuzione                               0.44391164
LogPremio                                    -0.48927304
```

Fig. 33: MIX regression estimated parameters

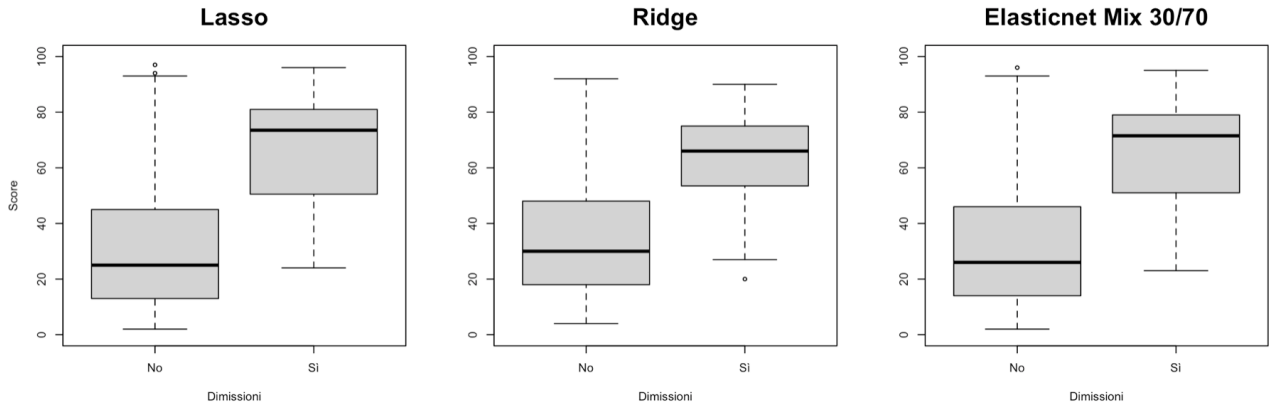


Fig. 34: distribution of the Score on the test set of the three models

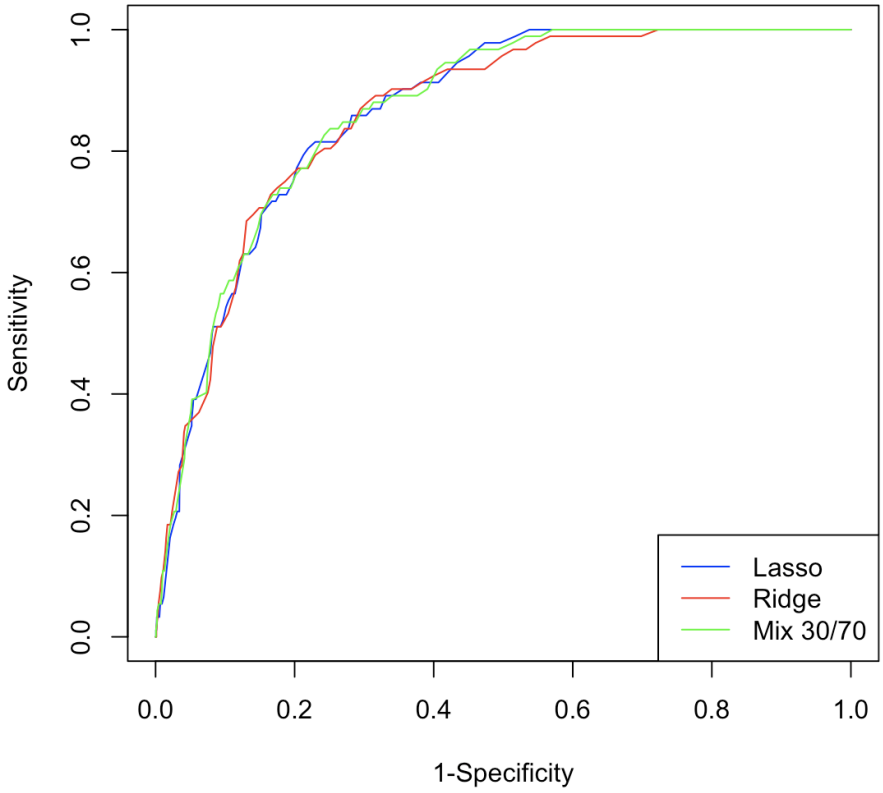


Fig. 35: ROC curve of the three models

```
Lasso
```

```
Confusion Matrix and Statistics

          Reference
Prediction   No   Sì
        No 1305   21
        Sì  332   71

               Accuracy : 0.7958
                 95% CI : (0.7761, 0.8146)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 1

                  Kappa : 0.2192

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.77174
            Specificity : 0.79719
         Pos Pred Value : 0.17618
         Neg Pred Value : 0.98416
             Prevalence : 0.05321
         Detection Rate : 0.04106
   Detection Prevalence : 0.23308
      Balanced Accuracy : 0.78446

       'Positive' Class : Sì
```

Fig. 36: LASSO Confusion Matrix and Statistics

```
Ridge
```

```
Confusion Matrix and Statistics

          Reference
Prediction   No   Sì
        No 1261   19
        Sì  376   73

               Accuracy : 0.7715
                 95% CI : (0.751, 0.7911)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1991

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.79348
            Specificity : 0.77031
         Pos Pred Value : 0.16258
         Neg Pred Value : 0.98516
             Prevalence : 0.05321
         Detection Rate : 0.04222
   Detection Prevalence : 0.25969
      Balanced Accuracy : 0.78189

       'Positive' Class : Sì
```

Fig. 37: RIDGE Confusion Matrix and Statistics

```
Elasticnet Mix 30/70
```

```
Confusion Matrix and Statistics

          Reference
Prediction   No   Sì
        No 1294   21
        Sì  343   71

               Accuracy : 0.7895
                 95% CI : (0.7695, 0.8085)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 1

                  Kappa : 0.212

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.77174
            Specificity : 0.79047
         Pos Pred Value : 0.17150
         Neg Pred Value : 0.98403
             Prevalence : 0.05321
         Detection Rate : 0.04106
   Detection Prevalence : 0.23944
      Balanced Accuracy : 0.78110

       'Positive' Class : Sì
```

Fig. 38: ELASTICNET Confusion Matrix and Statistics

4.2.2 Naïve Bayes

We have defined the Laplace estimator, which serves to help the algorithm make a better classification in those cases in which "an event never occurs for one or more levels of the class and therefore their joint probability is zero" (Lantz, 2019).
"The Laplace estimator adds a small number to each of the counts in the frequency table, which ensures that each feature has a non-zero probability of occurring with each class" (Lantz, 2019).
We used 10-fold cross-validation to set Laplace factor correction (fL).
The target label was Dimissioni = "Sì" (Leave).
We tried both 80:20 and 70:30 train-test split.

The results for the two different divisions of the training and the test set are shown below.

80:20 split

```
Naive Bayes

4615 samples
  14 predictor
   2 classes: 'Leave', 'Stay'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 4153, 4154, 4153, 4154, 4153, 4153, ...
Resampling results across tuning parameters:

  fL  ROC        Sens       Spec
   0  0.8420343  0.7961055  0.7512306
   1  0.8421171  0.7961055  0.7520692
   2  0.8417149  0.7970064  0.7529078
   3  0.8414164  0.7965580  0.7537446
   4  0.8411270  0.7974589  0.7520710
   5  0.8409205  0.7970105  0.7524894
   6  0.8407514  0.7983558  0.7524894
   7  0.8405599  0.7965580  0.7520692
   8  0.8403304  0.7965580  0.7524876
   9  0.8401574  0.7965580  0.7520692
  10  0.8399392  0.7965580  0.7524876

Tuning parameter 'usekernel' was held constant at a value of FALSE

Tuning parameter 'adjust' was held constant at a value of 0
ROC was used to select the optimal model using the largest value.
The final values used for the model were fL = 1, usekernel = FALSE and a
djust
 = 0.
```

Fig. 39: 10-fold validation for Laplace factor on the training set

From the figure we can see that the algorithm chooses the value 1 for the Laplace factor. We have, however, decided to choose the value 0 for the factor.
This decision is due to the fact that the results with the value 0 and the value 1 for the Laplace factor are very similar.

We therefore trained the Naïve Bayes model with the fL parameter set to 0 and then we calculated the probability of resigning on the test set.

The variable "Score" shows the probability that a worker will resign or not (where 100 is an extreme probability of resigning). The graph below represents the distribution of the “Score” calculated on the test set. We observe that the values "Sì" of resignation are quite separate from the values "No".

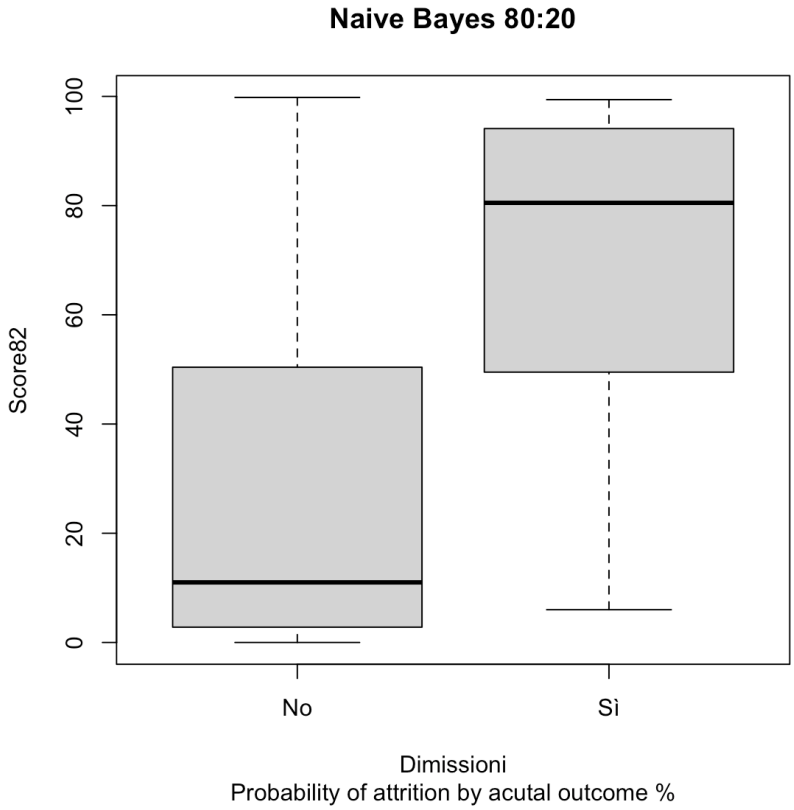


Fig. 40: distribution of the Score on the test set

We then calculated the area under the curve, the confusion matrix and other performance measures. The results are shown in the Fig. 14.
We note, from the value of sensitivity and specificity, that the algorithm correctly identifies 74% of employees who resign and 75% of workers who do not resign.

```
Confusion Matrix and Statistics

          Reference
Prediction  No  SÃ¬
        No 817  16
        SÃ¬ 274  45

               Accuracy : 0.7483
                 95% CI : (0.7222, 0.7731)
    No Information Rate : 0.947
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1624

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.73770
            Specificity : 0.74885
         Pos Pred Value : 0.14107
         Neg Pred Value : 0.98079
             Prevalence : 0.05295
         Detection Rate : 0.03906
   Detection Prevalence : 0.27691
      Balanced Accuracy : 0.74328

       'Positive' Class : SÃ¬
```

Fig. 41: Naïve Bayes Confusion Matrix and Statistics

We did the same calculation with a 70/30 split of the training set and the test set, in order to have more cases in the test set, and we saw that the results didn't change much.

70:30 split

```
Naive Bayes

4615 samples
  14 predictor
   2 classes: 'Leave', 'Stay'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 4154, 4153, 4153, 4154, 4154, 4154, ...
Resampling results across tuning parameters:

  fL  ROC        Sens       Spec
   0  0.8415487  0.7979336  0.7521061
   1  0.8415398  0.7992789  0.7525245
   2  0.8412301  0.7992789  0.7525245
   3  0.8409952  0.8001757  0.7529429
   4  0.8408657  0.8010766  0.7525245
   5  0.8407889  0.8006282  0.7525245
   6  0.8406481  0.8001778  0.7521061
   7  0.8404375  0.7997293  0.7529447
   8  0.8401950  0.8001778  0.7525263
   9  0.8400014  0.7997273  0.7529447
  10  0.8397531  0.7997273  0.7525263

Tuning parameter 'usekernel' was held constant at a value of FALSE

Tuning parameter 'adjust' was held constant at a value of 0
ROC was used to select the optimal model using the largest value.
The final values used for the model were fL = 0, usekernel = FALSE and a
djust
 = 0.
```

Fig. 42: 10-fold validation for Laplace factor

In this case, the algorithm itself sets the Laplace factor's value to zero.
The figures below represent the distribution of the "Score" and the confusion matrix. Fig. 16 shows that the score values with the 70:30 split are more distinct than with 80:20. From the Fig. 17 we can see that the sensitivity with the 70:30 slit is 80%, which is a good result.

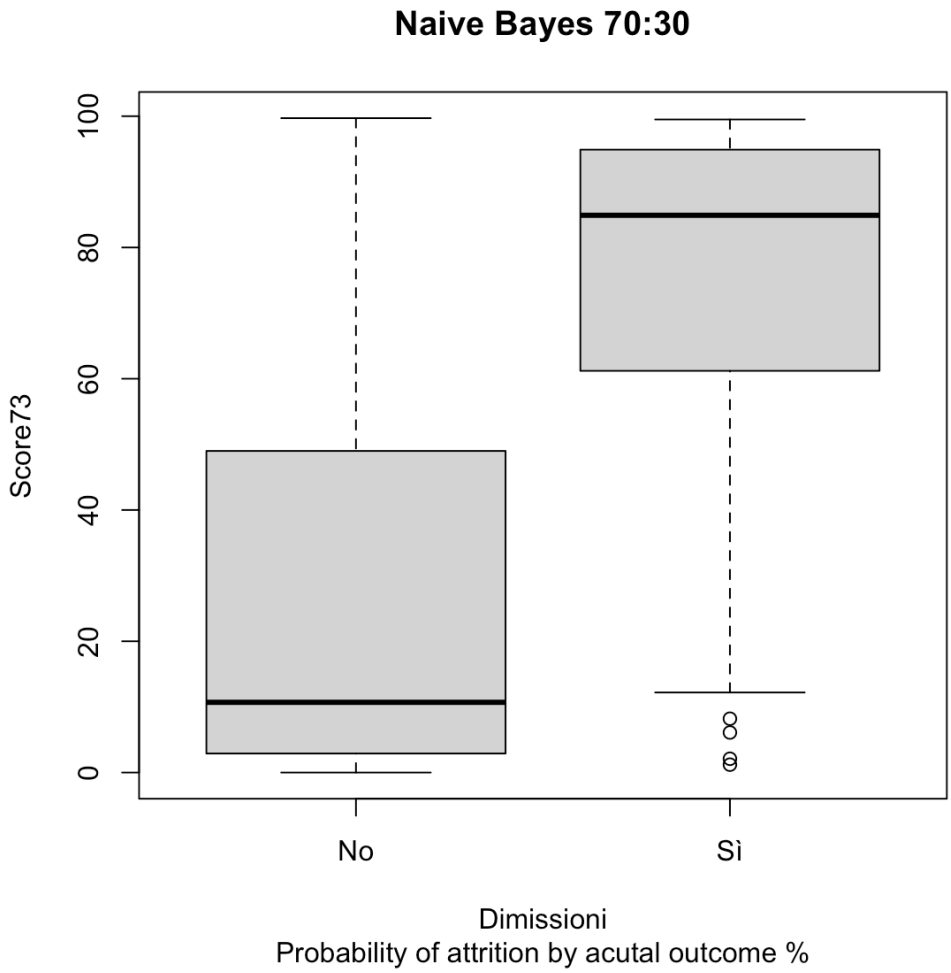


Fig. 43: distribution of the Score on the test set

```
Confusion Matrix and Statistics

          Reference
Prediction   No   SÃ¬
        No 1234   18
        SÃ¬  403   74

               Accuracy : 0.7565
                 95% CI : (0.7356, 0.7766)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1876

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.80435
            Specificity : 0.75382
         Pos Pred Value : 0.15514
         Neg Pred Value : 0.98562
             Prevalence : 0.05321
         Detection Rate : 0.04280
   Detection Prevalence : 0.27588
      Balanced Accuracy : 0.77908

       'Positive' Class : SÃ¬
```

Fig. 44: Naïve Bayes Confusion Matrix and Statistics

Since there is not a variable-importance method for Naïve Bayes in *caret*, we calculated the probability of the characteristics based on the model based on whether employees resign or not, as follows:

- For qualitative features with values $v_1, v_2, \ldots, v_p$:

$$\Pr\{\text{Leave}|\text{Feature} = v_j\}\ ,\ j = 1, 2, \ldots, p$$

- For quantitative features:

$$\text{Feature} \sim N(\mu, \sigma) \text{ column [,1] is } \mu \text{ and [,2] is } \sigma$$

The results are shown in the Figure 18 and show that the variables that most affect the probability to resign are: “Genere”, “Posizione”, “Orario”, “Grado”, “Retention”, “ClasseEta”, “ClasseAnniServizio”, “Job_description”, “LogPremio”. Since in the Naïve Bayes algorithm it is assumed that all characteristics are independent of each other, the results do not show the interdependence between the characteristics.

```
$Genere
        var
grouping   Femmina   Maschio
      No 0.3418025 0.6581975
      SÃ¬ 0.2996926 0.7003074

$ClasseStruttura
        var
grouping      Centro        Rete     SocietÃ
      No 0.325023969 0.669702780 0.005273250
      SÃ¬ 0.318135246 0.675204918 0.006659836

$Posizione
        var
grouping    Addetto Responsabile
      No 0.85953979   0.14046021
      SÃ¬ 0.94518443   0.05481557

$Orario
        var
grouping  Full Time  Part Time
      No 0.95397891 0.04602109
      SÃ¬ 0.97028689 0.02971311

$Grado
        var
grouping          1          2          3          4          5
      No 0.28906999 0.26366251 0.25215724 0.14621285 0.04889741
      SÃ¬ 0.41547131 0.23463115 0.25461066 0.05379098 0.04149590

$Contratto
        var
grouping      Altro    Credito  Dirigenti Funzionari
      No 0.06807287 0.69079578 0.04889741 0.19223394
      SÃ¬ 0.06045082 0.72028689 0.04149590 0.17776639
```

```
$Talento
        var
grouping         No         SÃ¬
      No 0.94870566 0.05129434
      SÃ¬ 0.93954918 0.06045082

$Gratifica
        var
grouping         No         SÃ¬
      No 0.96548418 0.03451582
      SÃ¬ 0.97387295 0.02612705

$Retention
        var
grouping          No          SÃ¬
      No 0.998082454 0.001917546
      SÃ¬ 0.966188525 0.033811475

$ClasseEta
        var
grouping    (0, 25]   (25, 30]   (30, 40]   (40, 50] (50, +Inf]
      No 0.02780441 0.23489933 0.23777565 0.29434324 0.20517737
      SÃ¬ 0.09067623 0.44774590 0.25256148 0.15215164 0.05686475

$ClasseAnniServizio
        var
grouping      [0,3]     (3,10]    (10,20]    (20,30]   (30,Inf]
      No 0.15196548 0.14908917 0.37392138 0.20949185 0.11553212
      SÃ¬ 0.38883197 0.32838115 0.22387295 0.04508197 0.01383197

$Job_description
        var
grouping        ALTRO AREA CREDITI AREA MERCATI AREA TECNICA E LOGISTICA
      No 0.0268456376 0.0364333653 0.0004793864             0.0182166826
      SÃ¬ 0.0399590164 0.0025614754 0.0148565574              0.0169057377
```

```
grouping AUDIT E COMPLIANCE BACK OFFICE E AMMINISTRAZIONE
      No       0.0177372963                  0.0201342282
      SÃ¬       0.0215163934                  0.0210040984
        var
grouping CONSULENZA FINANZIARIA CORPORATE BANKING CUSTOMER ASSISTANCE
      No           0.1438159156      0.0570469799        0.0752636625
      SÃ¬           0.2612704918      0.0343237705        0.0517418033
        var
grouping    DIREZIONE      FINANZA GOVERNANCE E ORGANIZZAZIONE          ICT
      No 0.0057526366 0.0134228188                0.0206136146 0.0306807287
      SÃ¬ 0.0153688525 0.0128073770                0.0189549180 0.0373975410
        var
grouping    INSURANCE        LEGAL PEOPLE MANAGEMENT RETAIL BANKING
      No 0.0009587728 0.0105465005      0.0143815916   0.2708533078
      SÃ¬ 0.0000000000 0.0235655738      0.0020491803   0.1711065574
        var
grouping RISK E AL MANAGEMENT SERVIZI CONDIVISI SERVIZI DI CASSA
      No         0.0129434324      0.0580057526     0.1452540748
      SÃ¬         0.0327868852      0.0092213115     0.1577868852
        var
grouping SICUREZZA E PREVENZIONE STAFF DI DIREZIONE SVILUPPO ASSET
      No            0.0009587728       0.0004793864   0.0062320230
      SÃ¬            0.0000000000       0.0000000000   0.0220286885
        var
grouping SVILUPPO CLIENTI
      No     0.0129434324
      SÃ¬     0.0327868852

$LogRetribuzione
       [,1]      [,2]
No 3.596948 0.1621317
SÃ¬ 3.549860 0.1920703

$LogPremio
       [,1]     [,2]
No 2.942136 1.335374
SÃ¬ 1.432431 1.732942
```

Fig. 45: probability to resign based on characteristics

4.2.3 Decision Tree

Decision Tree Classifiers has been implemented on the training set.
We made four graphs: one with *prp* that is a function that returns a more synthetic chart, and three with *rpart.plot* that returns a more detailed chart. *Rpart.plot* was applied first on resigned, and then on both cases, to see in detal the difference between employee who resigned and those who don't. The display of the *prp* function has four constituents: the node labels, the split labels, the branch lines, and the optional node numbers (Milborrow, 2020).
In the case of Figures 20 and 21, each node shows: the predicted class (Not resigned, "No", and resigned, "Sì"), (ii) the predicted probability of remain in the organization (iii) the percentage of observations in the node (Milborrow, 2020).
In the case of Figures 17 and 19, however, where only the resigned class were classified in the model, each node shows: (i) the predicted value and (ii) the percentage of observations in the node.
The figures below illustrate the Decision Tree and the rules.

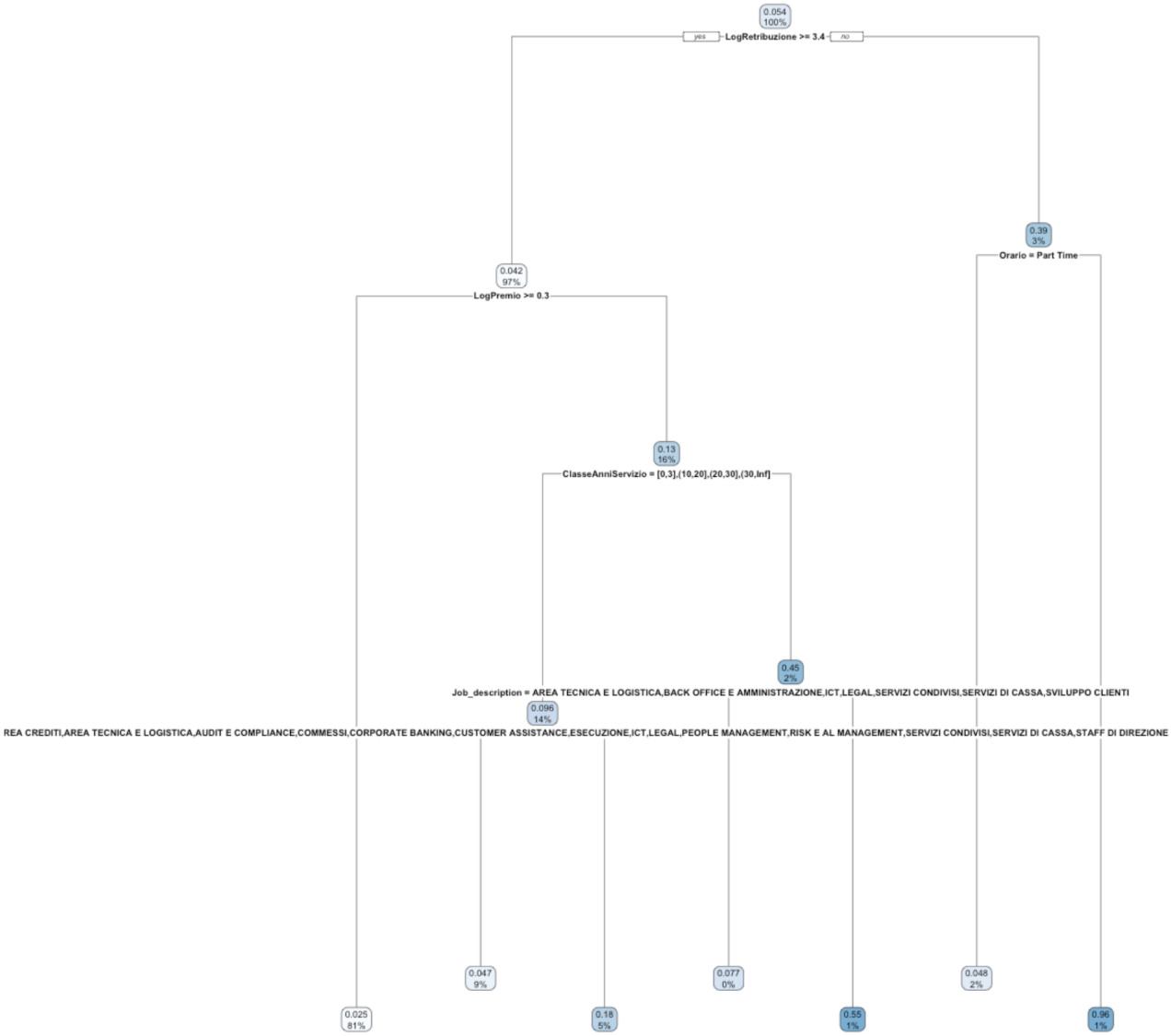


Fig. 46: Decision Tree for resigned made with rpart.plot function

```
 Dimissioni == "Sì"
cover
             0.025 when LogRetribuzione >= 3.4 & LogPremio >= 0.3
81%
             0.047 when LogRetribuzione >= 3.4 & LogPremio <  0.3 & [0,3] or (10,20] or (20,30] or (30,Inf] & AR
EA CREDITI or AREA TECNICA E LOGISTICA or AUDIT E COMPLIANCE or COMMESSI or CORPORATE BANKING or CUSTOMER ASSISTA
NCE or ESECUZIONE or ICT or LEGAL or PEOPLE MANAGEMENT or RISK E AL MANAGEMENT or SERVIZI CONDIVISI or SERVIZI DI
CASSA or STAFF DI DIREZIONE                  9%
             0.048 when LogRetribuzione <  3.4
& Part Time      2%
             0.077 when LogRetribuzione >= 3.4 & LogPremio <  0.3 &                                  (3,10] &
AREA TECNICA E LOGISTICA or BACK OFFICE E AMMINISTRAZIONE or ICT or LEGAL or SERVIZI CONDIVISI or SERVIZI DI CASS
A or SVILUPPO CLIENTI                 0%
             0.180 when LogRetribuzione >= 3.4 & LogPremio <  0.3 & [0,3] or (10,20] or (20,30] or (30,Inf] &
ALTRO or AREA MERCATI or BACK OFFICE E AMMINISTRAZIONE or CONSULENZA FINANZIARIA or FINANZA or GOVERNANCE E ORGAN
IZZAZIONE or RETAIL BANKING or SVILUPPO ASSET or SVILUPPO CLIENTI                 5%
             0.547 when LogRetribuzione >= 3.4 & LogPremio <  0.3 &                                  (3,10] &
CONSULENZA FINANZIARIA or CORPORATE BANKING or CUSTOMER ASSISTANCE or DIREZIONE or GOVERNANCE E ORGANIZZAZIONE or
RETAIL BANKING or SVILUPPO ASSET                 1%
             0.959 when LogRetribuzione <  3.4
& Full Time      1%
```

Fig. 47: detail of the rules of the Decision Tree shown in Fig. 12

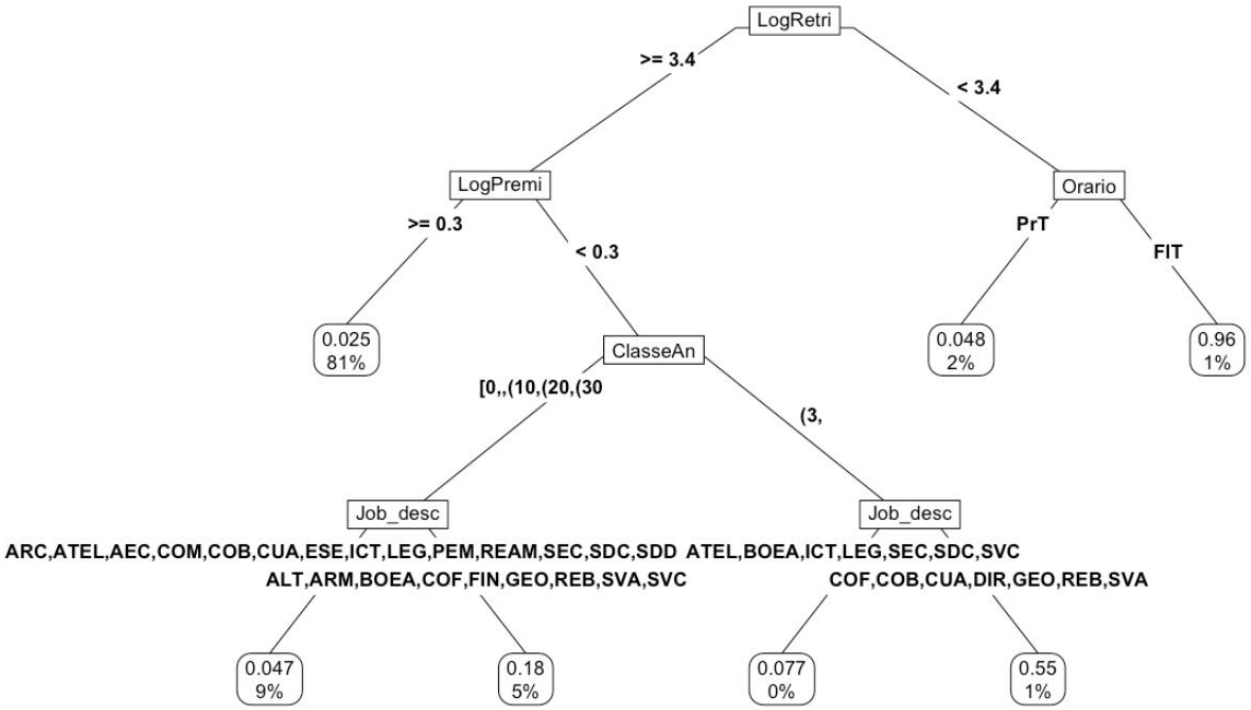


Fig. 48: Decision Tree for resigned made with prp function

Fig. 17 and Fig. 19 are two different visualizations that shows the same tree that only predicts cases of resignation. Fig. 17 shows the Decision Tree for the resignation cases made with the *rpart* function, while Fig. 19 shows the same tree build with the *prp* function, that automatically scales and adjusts the displayed tree for best fit (Milborrow, 2020), making the tree clearer and readable.

Starting from the right, the rules and the graph show that when workers earn a not much (when the average salary, measured with the logarithm, is below 3.4), to have a part time or full time affects the decision to leave the company or not. The rightmost node shows that full-time employees with a low average wage have a 0.96 probability of leave the organization, and 1% of the employees fell into this rule. This result also intuitively makes sense, because part-time workers are expected not to look forward to very high wages and are less interested in making a career.

In the case of workers who earn much, on the other hand, the variables bonus and the tenure are important. Employees who receive a bonus (measured with the logarithm) above 0.3 have a 0.025 probability of leave the organization (very low, that is it is very unlikely that they will leave the organization), and 81% of employees belong to this class. For workers who receive a bonus below 0.3, tenure and job description are the variables that make the difference. The employees who have a tenure that belong to the class (3,10] (a middle way therefore between newcomers and employee with high tenure) and the other are distinguished. The tree shows that workers keep belonging to class (3, 10] and to the job description: “CONSULENZA FINANZIARIA or CORPORATE BANKING or CUSTOMER ASSISTANCE or DIREZIONE or GOVERNANCE E ORGANIZZAZIONE or RETAIL BANKING or SVILUPPO ASSET” are more likely to leave the organization (probability of 0.55, 1% of the employees).
These results make sense because it is intuitive to think that employees who belong to the functions listed above, where there is a good demand in the labour market, who are at a stage where they have made a good career, seek to resell themselves in the labour market, and in cases where they

have a low bonus this is more likely to happen. For example, those employee who work in "Consulenza finanziaria" have already made a portfolio of customers, and then can negotiate with other organizations more attractive working conditions, because they carry their own client portfolio.

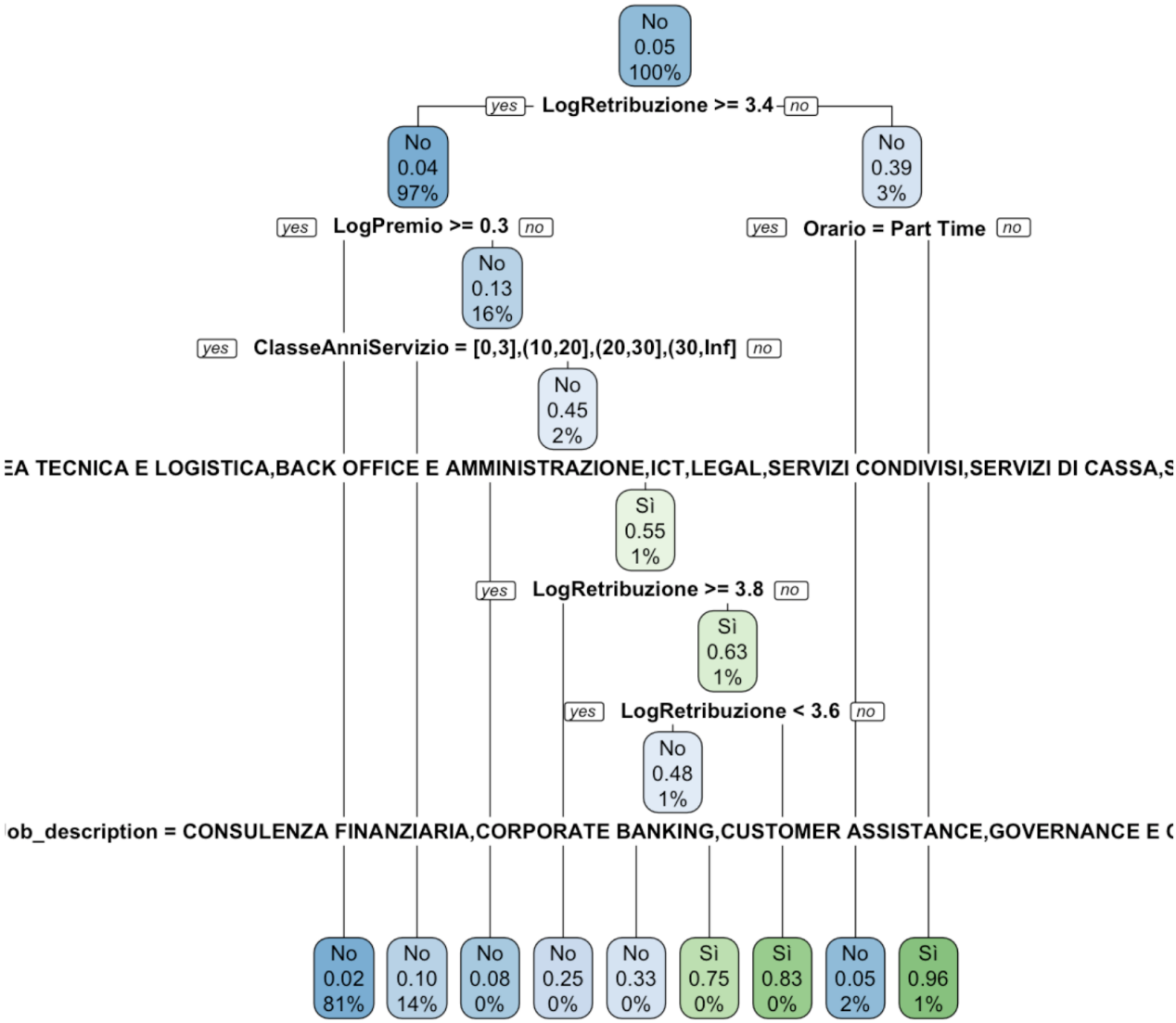


Fig. 49: Decision Tree for the two class (Resigned and not) made with rpart.plot function

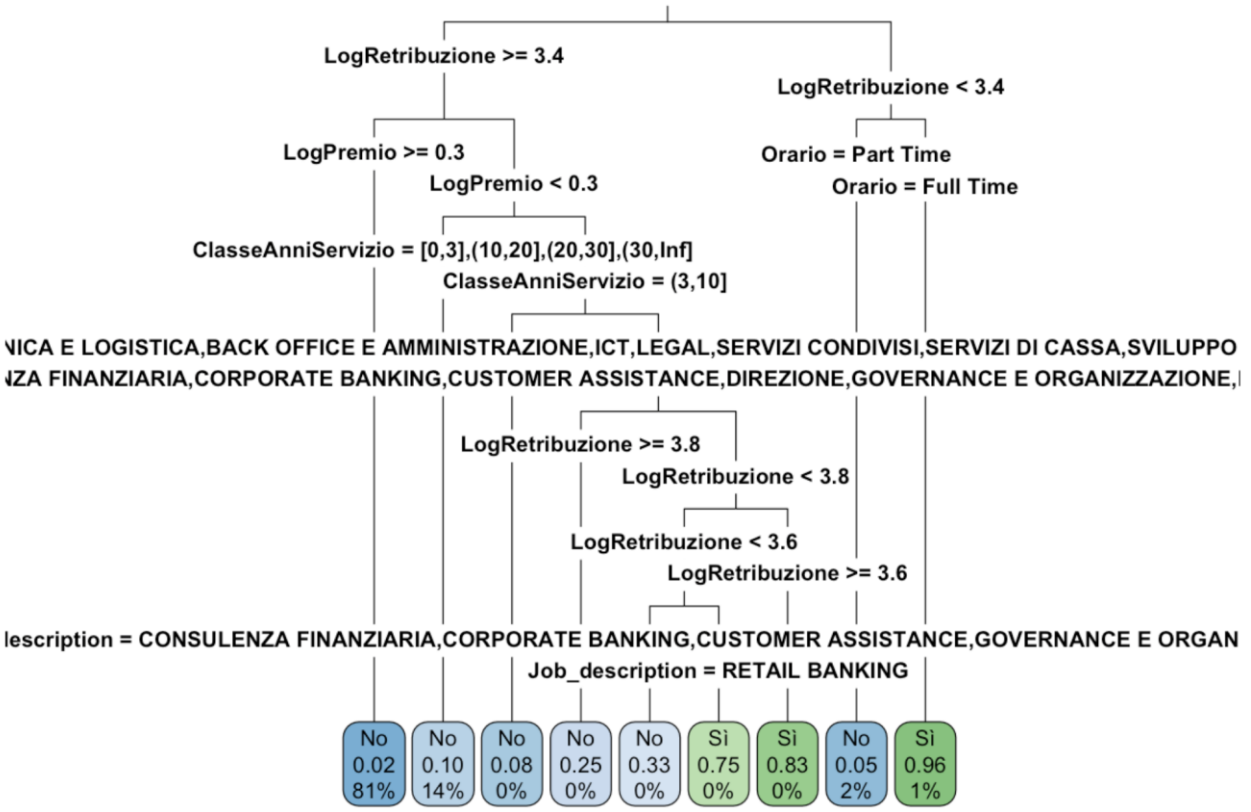


Fig. 50: Decision Tree for the two class (Resigned and not) made with rpart.plot function (a clearer visualization)

```
 Dimissioni
cover
       0.02 when LogRetribuzione >=        3.4 & LogPremio >= 0.3
81%
       0.05 when LogRetribuzione <  3.4
& Part Time      2%
       0.08 when LogRetribuzione >=        3.4 & LogPremio <  0.3 &                                (3,10] &
AREA TECNICA E LOGISTICA or BACK OFFICE E AMMINISTRAZIONE or ICT or LEGAL or SERVIZI CONDIVISI or SERVIZI DI CASS
A or SVILUPPO CLIENTI                  0%
       0.10 when LogRetribuzione >=        3.4 & LogPremio <  0.3 & [0,3] or (10,20] or (20,30] or (30,Inf]
14%
       0.25 when LogRetribuzione >=        3.8 & LogPremio <  0.3 &                                (3,10] & CON
SULENZA FINANZIARIA or CORPORATE BANKING or CUSTOMER ASSISTANCE or DIREZIONE or GOVERNANCE E ORGANIZZAZIONE or RE
TAIL BANKING or SVILUPPO ASSET                 0%
       0.33 when LogRetribuzione is 3.4 to 3.6 & LogPremio <  0.3 &                                (3,10] &
CONSULENZA FINANZIARIA or CORPORATE BANKING or CUSTOMER ASSISTANCE or GOVERNANCE E ORGANIZZAZIONE
0%
       0.75 when LogRetribuzione is 3.4 to 3.6 & LogPremio <  0.3 &                                (3,10] &
RETAIL BANKING                 0%
       0.83 when LogRetribuzione is 3.6 to 3.8 & LogPremio <  0.3 &                                (3,10] & CON
SULENZA FINANZIARIA or CORPORATE BANKING or CUSTOMER ASSISTANCE or DIREZIONE or GOVERNANCE E ORGANIZZAZIONE or RE
TAIL BANKING or SVILUPPO ASSET                 0%
       0.96 when LogRetribuzione <  3.4
& Full Time      1%
```

Fig. 51: detail of the rules of the Decision Tree shown in Fig. 15

Fig. 20 and Fig. 21 are two different visualizations that shows the same tree that predict both the cases of employee who resign and those who not. Fig. 20 explicitly labels both left and right hand branches of each split, while Fig. 21 labels on the interior nodes (Milborrow, 2020). Fig. 17 explicits the relative rules.

The nodes colored in green gives the resignation cases, while the nodes colored in blue the remaining workers in the organization.

Starting at the bottom, the last rule says that full-time employees who receive an average wage (measured with the logarithm) below 3.4 have a probability of 0.96 to leave the organization. The last node number shows that only 1% of the employees fell into this rule. The other nodes representing the resignation cases of the Fig. 21 appear to show fictitious probabilities, because no case belongs to the node, or maybe there are very few cases, and the algorithm rounds the percentage to zero.
Two interior nodes of the Fig. 20 gives the resignation cases. The first says that employees with the job description that belong to one of the classes: “CONSULENZA FINANZIARIA or CORPORATE BANKING or CUSTOMER ASSISTANCE or GOVERNANCE E ORGANIZZAZIONE RETAIL BANKING”, who have a tenure that belong to the class (3,10] (a middle way therefore between newcomers and employee with high tenure), who have a bonus (measured with the logarithm) below 0.3 and who receive an average wage (measured with the logarithm) above 3.4, have a probability of 0.55 to leave the organization, and only 1% of the employees fell into this rule. The second green node following the one just described shows that employee that fall into the categories described above, and also receive an average wage (measured with the logarithm) below 3.8, have a probability of 0.63 to resign and 1% of the employees fell into this case.

Considering all of this evidence, it seems that the most important factors in predicting resignation are: remuneration, for low wages the type of work schedule (part-time or full-time), for higher wages the bonus, seniority and job description.

Based on the tree model, we calculated the probability of resigning on the test set.

The variable "Score" shows the probability that a worker will resign or not (where 100 is an extreme probability of resigning). The graph below represents the distribution of the “Score” calculated on the test set.

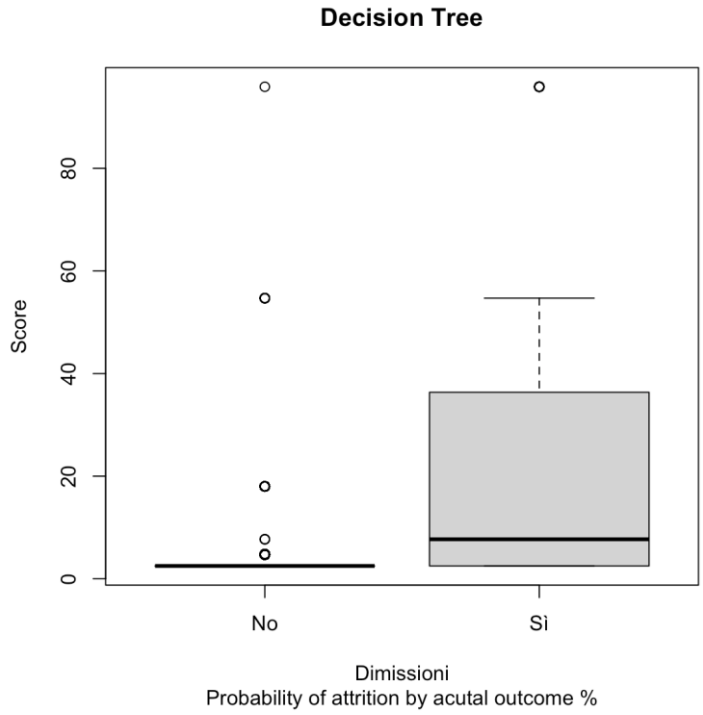


Fig. 52: distribution of the Score on the test set

We then calculated the ROC curve, the area under the curve, the confusion matrix and other performance measures. The results are shown in the Fig. 19 and Fig. 20. We can notice that the Decision Tree, compared to other models, has worse performance.

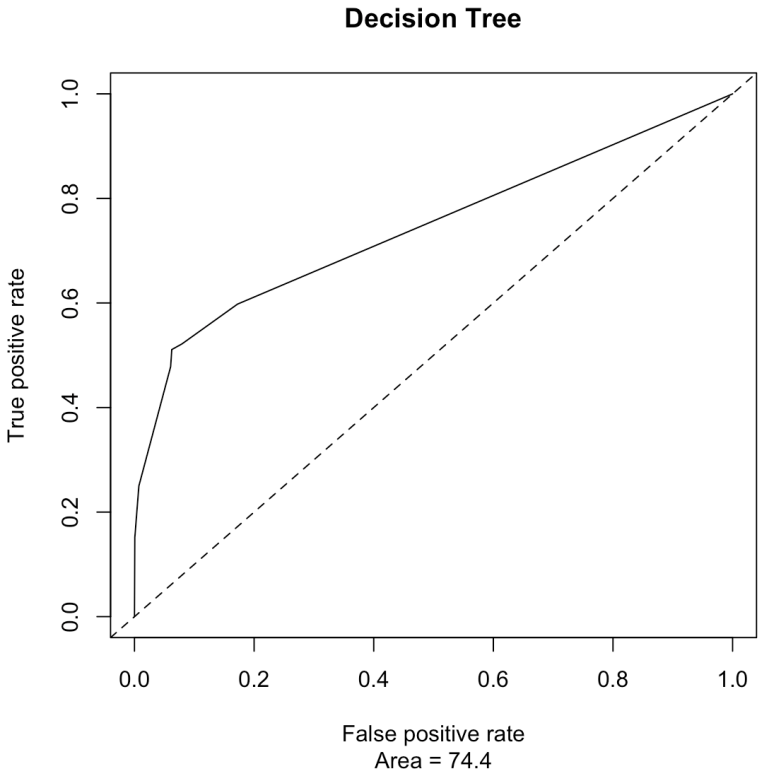


Fig. 53: Decision Tree's ROC curve

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  1535   45
     TRUE    102   47

               Accuracy : 0.915
                 95% CI : (0.9008, 0.9277)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 1

                  Kappa : 0.3471

 Mcnemar's Test P-Value : 3.86e-06

            Sensitivity : 0.9377
            Specificity : 0.5109
         Pos Pred Value : 0.9715
         Neg Pred Value : 0.3154
             Prevalence : 0.9468
         Detection Rate : 0.8878
   Detection Prevalence : 0.9138
      Balanced Accuracy : 0.7243

       'Positive' Class : FALSE
```

Fig. 54: Decision Tree's Confusion Matrix and Statistics

We then trained the model with the k-fold validation and calculated its performance measurements (Fig. 21).

```
CART

4038 samples
  14 predictor
   2 classes: 'No', 'Sì'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 3635, 3634, 3634, 3634, 3635, 3634, ...
Resampling results across tuning parameters:

  cp          ROC        Sens       Spec
  0.00000000  0.7717522  0.9926702  0.28593074
  0.01152074  0.6754591  0.9968586  0.23116883
  0.10368664  0.5311749  0.9997382  0.07337662

ROC was used to select the optimal model using the largest value.
The final value used for the model was cp = 0.
```

Fig. 55: Performance measures of the Decision Tree trained with k-fold validation

Finally, we assessed the importance of the individual variables in determining the result (Fig. 22). The decision tree is a model that has the advantage of being easily interpreted, and the importance of the variables is consistent with that of Figures 12, 14, 15 and 16, from which emerges the importance of predictors.

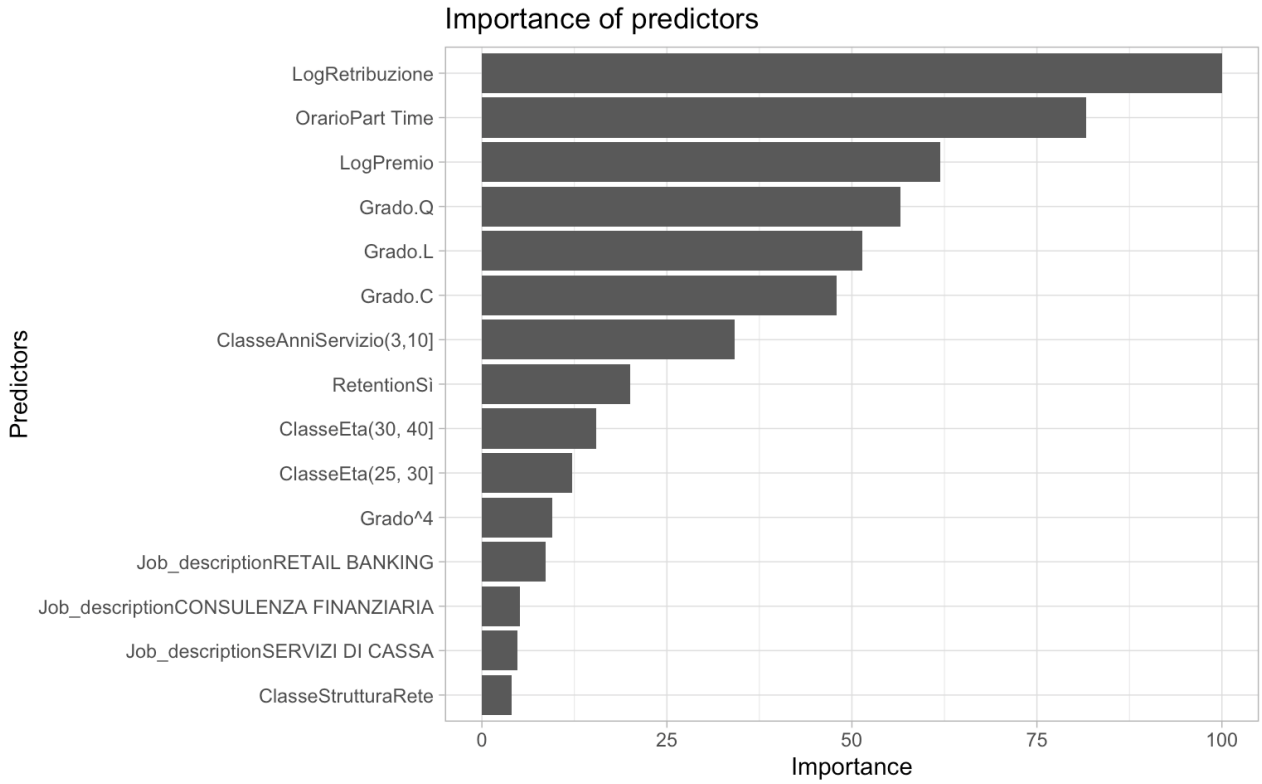


Fig. 56: Decision Tree's variable importance

Decision Tree with the ROSE correction

70:30 split

```
 Dimissioni == "Sì"
cover
              0.081 when LogPremio >= 1.4 &        (10,20] or (20,30] or (30,Inf] & AREA CREDITI or AREA TECNICA
E LOGISTICA or AUDIT E COMPLIANCE or BACK OFFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or
FINANZA or GOVERNANCE E ORGANIZZAZIONE or ICT or LEGAL or PEOPLE MANAGEMENT or RETAIL BANKING or SERVIZI CONDIVIS
I or SERVIZI DI CASSA or SICUREZZA E PREVENZIONE or STAFF DI DIREZIONE or SVILUPPO ASSET
29%
              0.179 when LogPremio <  1.4 &                              (30,Inf]
1%
              0.246 when LogPremio >= 1.4 &        (10,20] or (20,30] or (30,Inf] &
ALTRO or CONSULENZA FINANZIARIA or DIREZIONE or RISK E AL MANAGEMENT or SVILUPPO CLIENTI
&        Altro or Credito      6%
              0.350 when LogPremio >= 1.4 &                        [0,3] or (3,10] &
AREA CREDITI or BACK OFFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or FINANZA or INSURANCE
or PEOPLE MANAGEMENT or RETAIL BANKING or SERVIZI CONDIVISI or SERVIZI DI CASSA or SVILUPPO CLIENTI
11%
              0.474 when LogPremio <  1.4 & [0,3] or (3,10] or (10,20] or (20,30] &
AREA CREDITI or AREA TECNICA E LOGISTICA or CUSTOMER ASSISTANCE or ICT or LEGAL or RISK E AL MANAGEMENT or SERVIZ
I CONDIVISI or SERVIZI DI CASSA & LogRetribuzione >= 3.4                                  7%
              0.618 when LogPremio >= 1.4 &        (10,20] or (20,30] or (30,Inf] &
ALTRO or CONSULENZA FINANZIARIA or DIREZIONE or RISK E AL MANAGEMENT or SVILUPPO CLIENTI
& Dirigenti or Funzionari      4%
              0.709 when LogPremio >= 1.4 &                        [0,3] or (3,10] &
ALTRO or AREA MERCATI or AREA TECNICA E LOGISTICA or AUDIT E COMPLIANCE or CONSULENZA FINANZIARIA or DIREZIONE or
GOVERNANCE E ORGANIZZAZIONE or ICT or LEGAL or RISK E AL MANAGEMENT or SVILUPPO ASSET
14%
              0.840 when LogPremio <  1.4 & [0,3] or (3,10] or (10,20] or (20,30] &
ALTRO or AREA MERCATI or AUDIT E COMPLIANCE or BACK OFFICE E AMMINISTRAZIONE or CONSULENZA FINANZIARIA or CORPORA
TE BANKING or FINANZA or GOVERNANCE E ORGANIZZAZIONE or RETAIL BANKING or SVILUPPO ASSET or SVILUPPO CLIENTI & Lo
gRetribuzione >= 3.4                                  19%
              0.984 when LogPremio <  1.4 & [0,3] or (3,10] or (10,20] or (20,30]
& LogRetribuzione <  3.4                                    9%
```

Fig. 57: detail of the rules of the Decision Tree for resigned (not shown because the graph is unreadable)

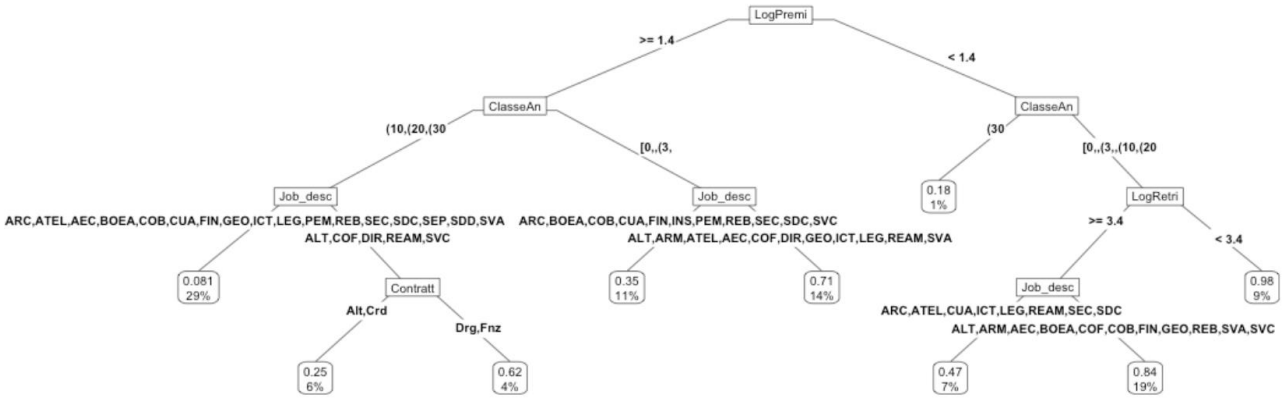


Fig. 58: Decision Tree for resigned made with prp function

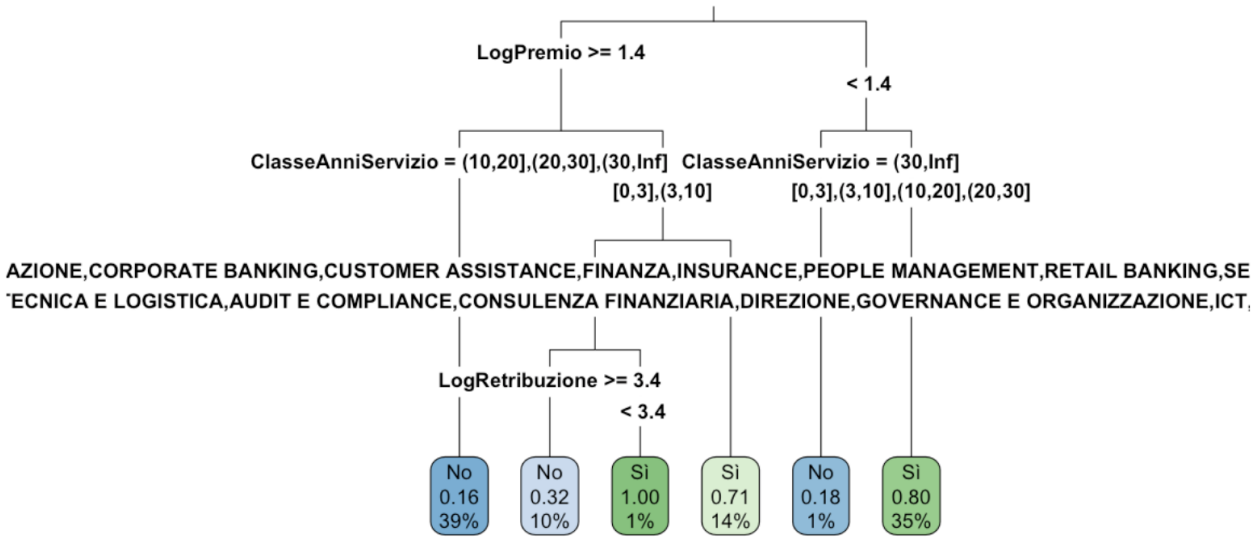


Fig. 59: Decision Tree for the two class (Resigned and not) made with rpart.plot function

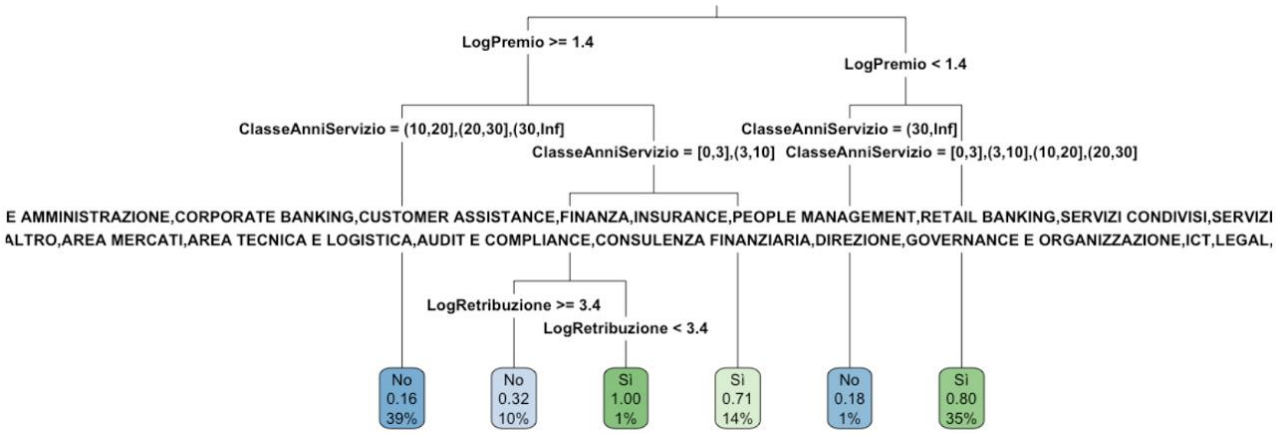


Fig. 60: Decision Tree for the two class (Resigned and not) made with rpart.plot function (a clearer visualization)

```
 Dimissioni
cover
       0.16 when LogPremio >= 1.4 &        (10,20] or (20,30] or (30,Inf]
39%
       0.18 when LogPremio <  1.4 &                                (30,Inf]
1%
       0.32 when LogPremio >= 1.4 &                      [0,3] or (3,10] & AREA CREDITI or BACK OFFICE E AMMINIS
TRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or FINANZA or INSURANCE or PEOPLE MANAGEMENT or RETAIL BANKI
NG or SERVIZI CONDIVISI or SERVIZI DI CASSA or SVILUPPO CLIENTI & LogRetribuzione >= 3.4     10%
       0.71 when LogPremio >= 1.4 &                      [0,3] or (3,10] &              ALTRO or AREA MERCATI o
r AREA TECNICA E LOGISTICA or AUDIT E COMPLIANCE or CONSULENZA FINANZIARIA or DIREZIONE or GOVERNANCE E ORGANIZZA
ZIONE or ICT or LEGAL or RISK E AL MANAGEMENT or SVILUPPO ASSET                                14%
       0.80 when LogPremio <  1.4 & [0,3] or (3,10] or (10,20] or (20,30]
35%
       1.00 when LogPremio >= 1.4 &                      [0,3] or (3,10] & AREA CREDITI or BACK OFFICE E AMMINIS
TRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or FINANZA or INSURANCE or PEOPLE MANAGEMENT or RETAIL BANKI
NG or SERVIZI CONDIVISI or SERVIZI DI CASSA or SVILUPPO CLIENTI & LogRetribuzione <  3.4      1%
```

Fig. 61: detail of the rules of the Decision Tree shown in Fig. 60 and 61

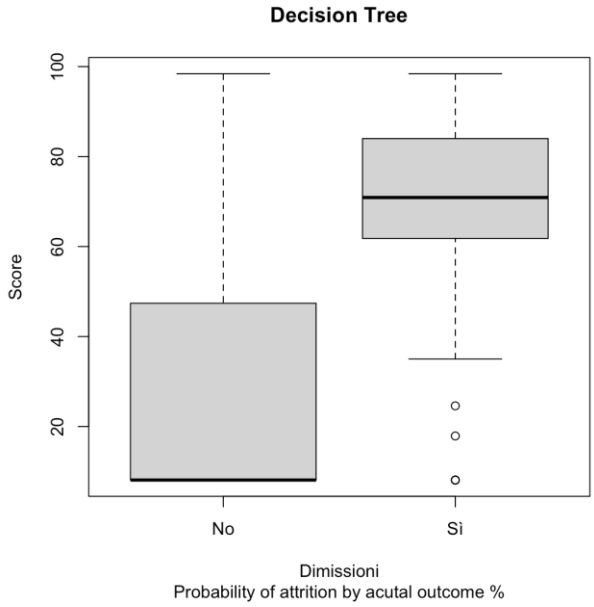


Fig. 62: distribution of the Score on the test set

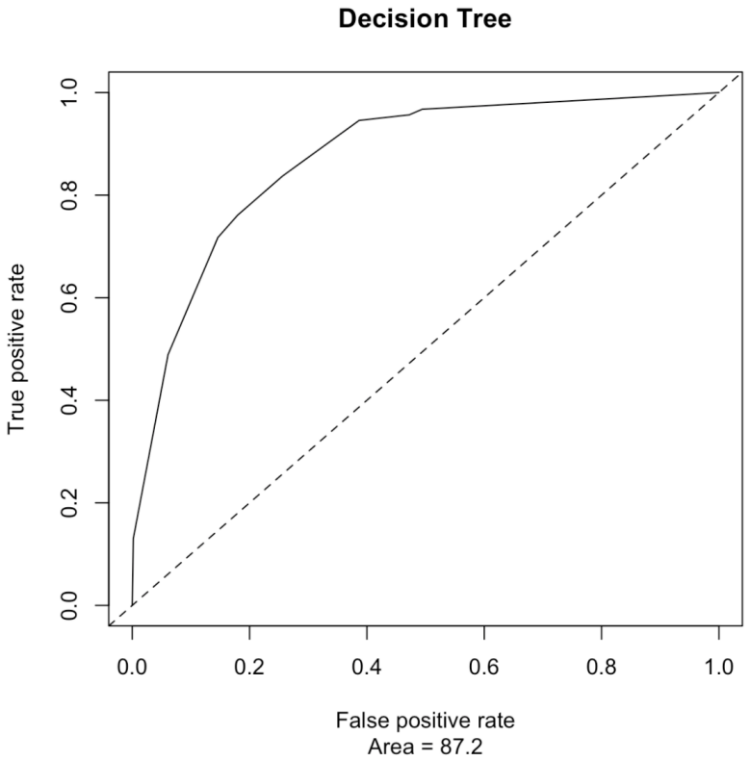


Fig. 63: Decision Tree's ROC curve

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  1343   22
     TRUE    294   70

               Accuracy : 0.8172
                 95% CI : (0.7982, 0.8352)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 1

                  Kappa : 0.2427

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.8204
            Specificity : 0.7609
         Pos Pred Value : 0.9839
         Neg Pred Value : 0.1923
             Prevalence : 0.9468
         Detection Rate : 0.7767
   Detection Prevalence : 0.7895
      Balanced Accuracy : 0.7906

       'Positive' Class : FALSE
```

Fig. 64: Decision Tree's Confusion Matrix and Statistics

```
CART

4038 samples
  14 predictor
   2 classes: 'No', 'Sì'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 3634, 3633, 3633, 3634, 3634, 3635, ...
Resampling results across tuning parameters:

  cp          ROC        Sens       Spec
  0.01656421  0.7702344  0.7808221  0.7080037
  0.01844262  0.7378336  0.8134868  0.6398796
  0.42366803  0.5605531  0.9462344  0.1748718

ROC was used to select the optimal model using the largest value.
The final value used for the model was cp = 0.01656421.
```

Fig. 65: Performance measures of the Decision Tree trained with k-fold validation

80:20 split

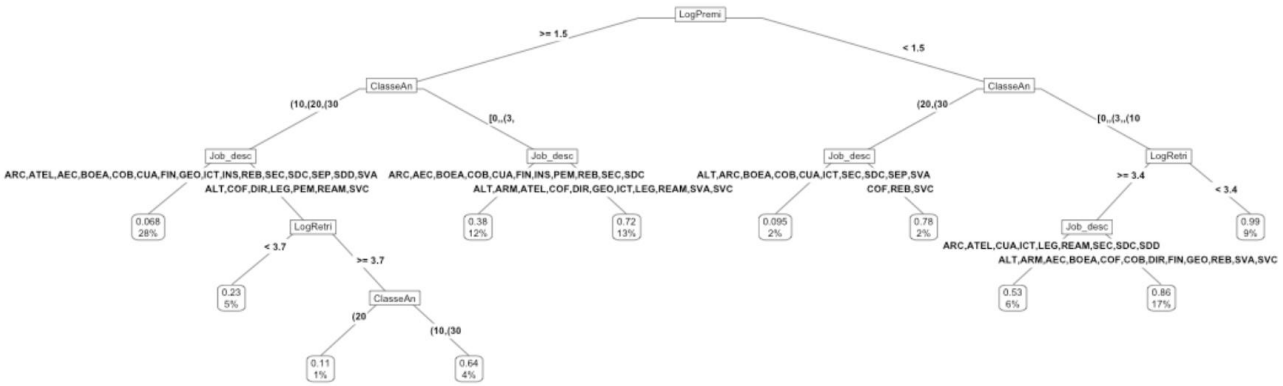


Fig. 66: Decision Tree for resigned made with prp function

```
 Dimissioni == "Sì"
cover
              0.068 when LogPremio >= 1.5 & (10,20] or (20,30] or (30,Inf] & AREA CREDITI or AREA TECNICA E LOGIS
TICA or AUDIT E COMPLIANCE or BACK OFFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or FINANZ
A or GOVERNANCE E ORGANIZZAZIONE or ICT or INSURANCE or RETAIL BANKING or SERVIZI CONDIVISI or SERVIZI DI CASSA o
r SICUREZZA E PREVENZIONE or STAFF DI DIREZIONE or SVILUPPO ASSET                              28%
              0.095 when LogPremio <  1.5 &            (20,30] or (30,Inf] &
ALTRO or AREA CREDITI or BACK OFFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or ICT or SERV
IZI CONDIVISI or SERVIZI DI CASSA or SICUREZZA E PREVENZIONE or SVILUPPO ASSET                                 2%
              0.111 when LogPremio >= 1.5 &                         (20,30] &
ALTRO or CONSULENZA FINANZIARIA or DIREZIONE or LEGAL or PEOPLE MANAGEMENT or RISK E AL MANAGEMENT or SVILUPPO CL
IENTI & LogRetribuzione >= 3.7      1%
              0.226 when LogPremio >= 1.5 & (10,20] or (20,30] or (30,Inf] &
ALTRO or CONSULENZA FINANZIARIA or DIREZIONE or LEGAL or PEOPLE MANAGEMENT or RISK E AL MANAGEMENT or SVILUPPO CL
IENTI & LogRetribuzione <  3.7      5%
              0.384 when LogPremio >= 1.5 &               [0,3] or (3,10] &
AREA CREDITI or AUDIT E COMPLIANCE or BACK OFFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE o
r FINANZA or INSURANCE or PEOPLE MANAGEMENT or RETAIL BANKING or SERVIZI CONDIVISI or SERVIZI DI CASSA
12%
              0.526 when LogPremio <  1.5 &     [0,3] or (3,10] or (10,20] &
AREA CREDITI or AREA TECNICA E LOGISTICA or CUSTOMER ASSISTANCE or ICT or LEGAL or RISK E AL MANAGEMENT or SERVIZ
I CONDIVISI or SERVIZI DI CASSA or STAFF DI DIREZIONE & LogRetribuzione >= 3.4      6%
              0.635 when LogPremio >= 1.5 &            (10,20] or (30,Inf] &
ALTRO or CONSULENZA FINANZIARIA or DIREZIONE or LEGAL or PEOPLE MANAGEMENT or RISK E AL MANAGEMENT or SVILUPPO CL
IENTI & LogRetribuzione >= 3.7      4%
              0.718 when LogPremio >= 1.5 &               [0,3] or (3,10] &
ALTRO or AREA MERCATI or AREA TECNICA E LOGISTICA or CONSULENZA FINANZIARIA or DIREZIONE or GOVERNANCE E ORGANIZZ
AZIONE or ICT or LEGAL or RISK E AL MANAGEMENT or SVILUPPO ASSET or SVILUPPO CLIENTI
13%
              0.779 when LogPremio <  1.5 &            (20,30] or (30,Inf] &
CONSULENZA FINANZIARIA or RETAIL BANKING or SVILUPPO CLIENTI                                  2%
              0.857 when LogPremio <  1.5 &     [0,3] or (3,10] or (10,20] &
ALTRO or AREA MERCATI or AUDIT E COMPLIANCE or BACK OFFICE E AMMINISTRAZIONE or CONSULENZA FINANZIARIA or CORPORA
TE BANKING or DIREZIONE or FINANZA or GOVERNANCE E ORGANIZZAZIONE or RETAIL BANKING or SVILUPPO ASSET or SVILUPPO
CLIENTI & LogRetribuzione >= 3.4     17%
              0.990 when LogPremio <  1.5 &     [0,3] or (3,10] or (10,20]
& LogRetribuzione <  3.4      9%
```

Fig. 67: detail of the rules of the Decision Tree shown in Fig. 67

```
 Dimissioni
cover
       0.09 when LogPremio <  1.5 &            (20,30] or (30,Inf] &                          ALTRO or AREA CREDIT
I or BACK OFFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or ICT or SERVIZI CONDIVISI or SER
VIZI DI CASSA or SICUREZZA E PREVENZIONE or SVILUPPO ASSET      2%
       0.16 when LogPremio >= 1.5 & (10,20] or (20,30] or (30,Inf]
39%
       0.38 when LogPremio >= 1.5 &               [0,3] or (3,10] & AREA CREDITI or AUDIT E COMPLIANCE or BACK O
FFICE E AMMINISTRAZIONE or CORPORATE BANKING or CUSTOMER ASSISTANCE or FINANZA or INSURANCE or PEOPLE MANAGEMENT
or RETAIL BANKING or SERVIZI CONDIVISI or SERVIZI DI CASSA     12%
       0.72 when LogPremio >= 1.5 &               [0,3] or (3,10] &                   ALTRO or AREA MERCATI or A
REA TECNICA E LOGISTICA or CONSULENZA FINANZIARIA or DIREZIONE or GOVERNANCE E ORGANIZZAZIONE or ICT or LEGAL or
RISK E AL MANAGEMENT or SVILUPPO ASSET or SVILUPPO CLIENTI     13%
       0.78 when LogPremio <  1.5 &            (20,30] or (30,Inf] &
CONSULENZA FINANZIARIA or RETAIL BANKING or SVILUPPO CLIENTI      2%
       0.83 when LogPremio <  1.5 &     [0,3] or (3,10] or (10,20]
32%
```

Fig. 68: detail of the rules of the Decision Tree for the two class (Resigned and not) (not shown because the graph is unreadable)

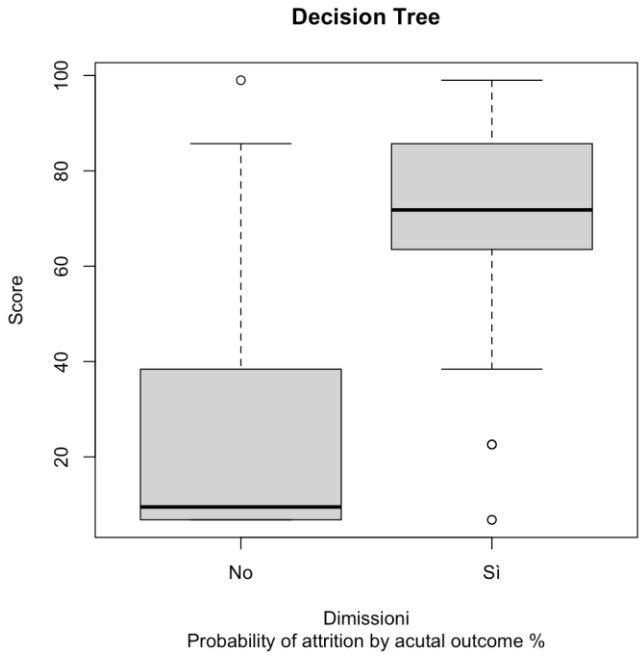


Fig. 69: distribution of the Score on the test set

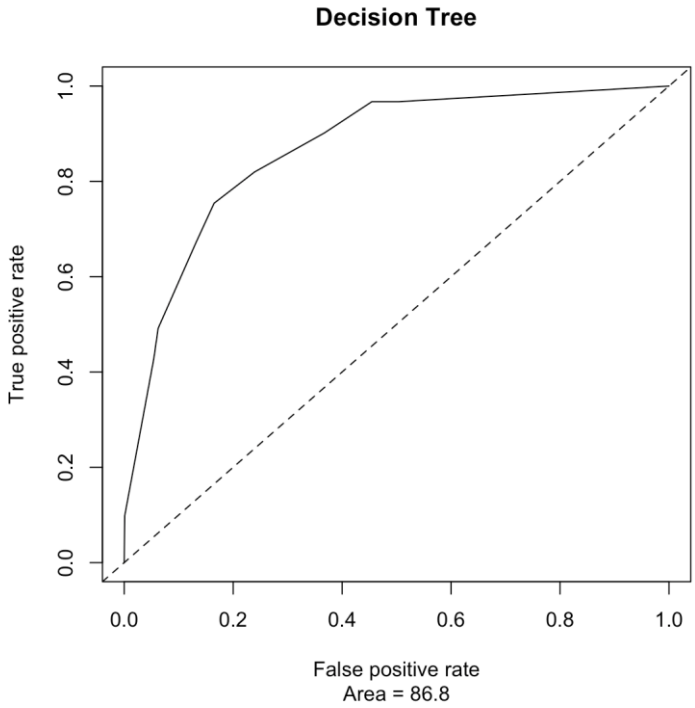


Fig. 70: Decision Tree's ROC curve

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE   830   11
     TRUE    261   50

               Accuracy : 0.7639
                 95% CI : (0.7383, 0.7881)
    No Information Rate : 0.947
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1978

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.7608
            Specificity : 0.8197
         Pos Pred Value : 0.9869
         Neg Pred Value : 0.1608
             Prevalence : 0.9470
         Detection Rate : 0.7205
   Detection Prevalence : 0.7300
      Balanced Accuracy : 0.7902

       'Positive' Class : FALSE
```

Fig. 71: Decision Tree's Confusion Matrix and Statistics

```
CART

4615 samples
  14 predictor
   2 classes: 'No', 'Sì'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 4153, 4153, 4153, 4154, 4153, 4154, ...
Resampling results across tuning parameters:

  cp          ROC        Sens       Spec
  0.01885945  0.7709783  0.7768204  0.7211611
  0.02267625  0.7470937  0.7818185  0.6780491
  0.42119443  0.6229824  0.9041841  0.3417808

ROC was used to select the optimal model using the largest value.
The final value used for the model was cp = 0.01885945.
```

Fig. 72: Performance measures of the Decision Tree trained with k-fold validation

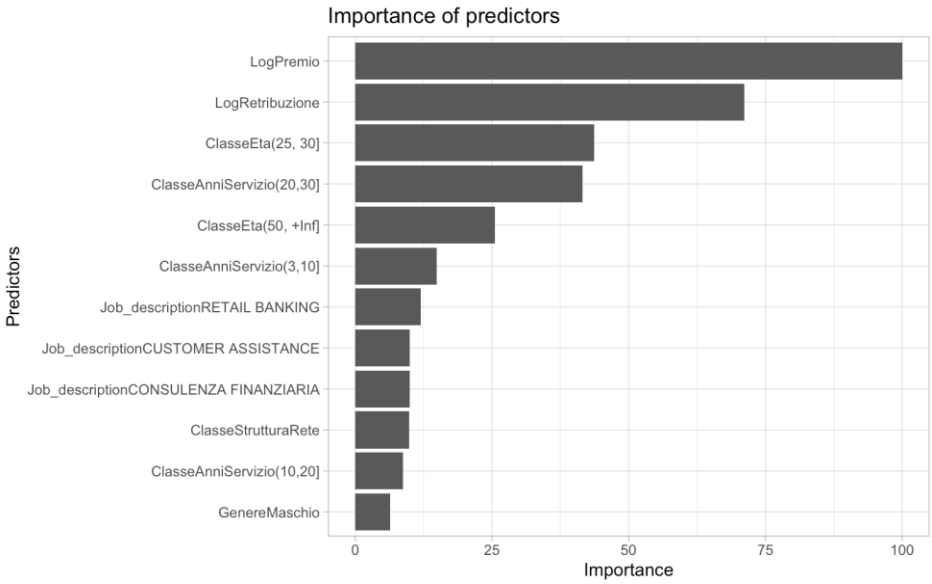


Fig. 73: Decision Tree's variable importance

4.2.4 Random Forest

70:30 split

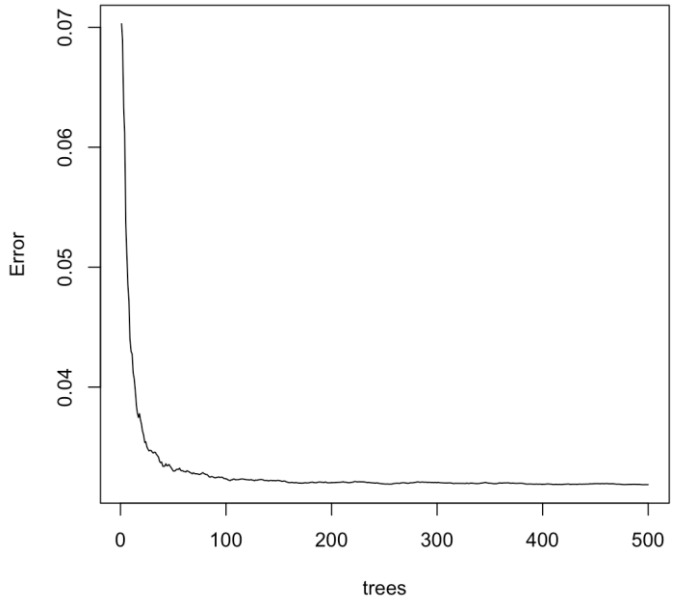


Fig. 74: error distribution as a function of the trees

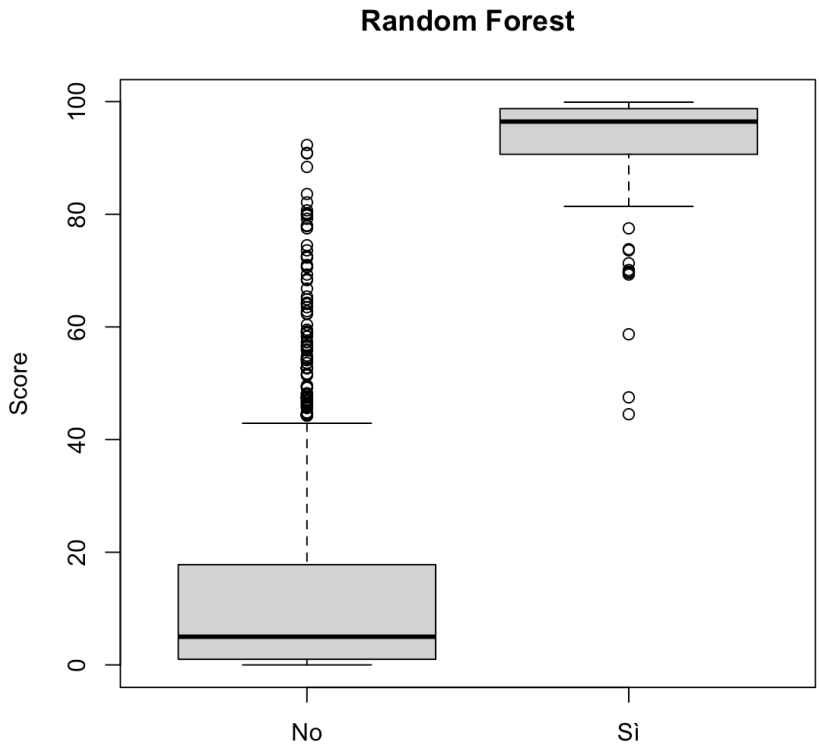


Fig. 75: distribution of the Score on the test set

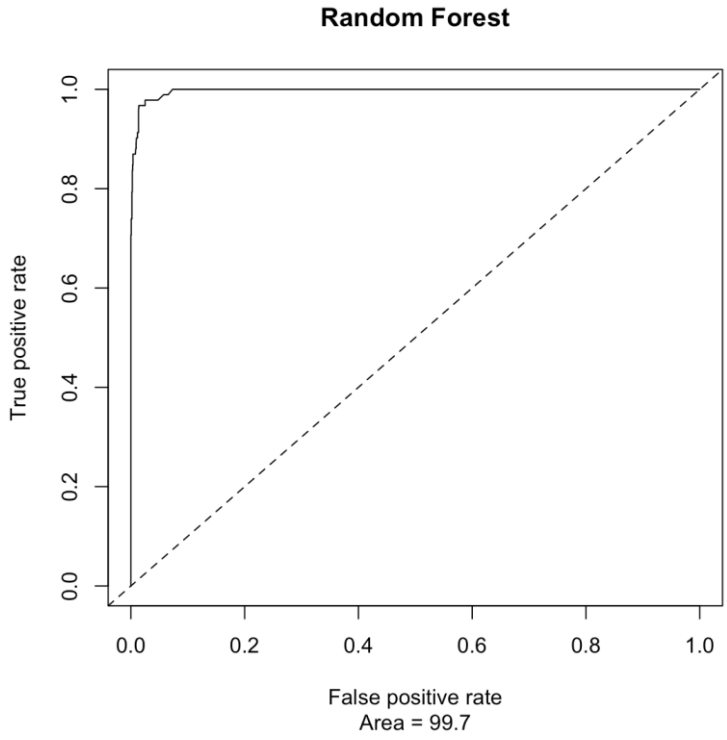


Fig. 76: Random Forest's ROC curve

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  1568    2
     TRUE     69   90

               Accuracy : 0.9589
                 95% CI : (0.9485, 0.9678)
    No Information Rate : 0.9468
    P-Value [Acc > NIR] : 0.01175

                  Kappa : 0.6967

 Mcnemar's Test P-Value : 4.773e-15

            Sensitivity : 0.9578
            Specificity : 0.9783
         Pos Pred Value : 0.9987
         Neg Pred Value : 0.5660
             Prevalence : 0.9468
         Detection Rate : 0.9069
   Detection Prevalence : 0.9080
      Balanced Accuracy : 0.9681

       'Positive' Class : FALSE
```

Fig. 77: Random Forest's Confusion Matrix and Statistics

```
Random Forest

4038 samples
  14 predictor
   2 classes: 'No', 'Sì'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 3634, 3635, 3633, 3634, 3634, 3634, ...
Resampling results across tuning parameters:

  mtry  ROC        Sens       Spec
   2    0.9092057  0.7948358  0.8637598
  25    0.9957804  0.9482172  0.9851491
  48    0.9944125  0.9376633  0.9851491

ROC was used to select the optimal model using the largest value.
The final value used for the model was mtry = 25.
```

Fig. 78: Performance measures of the Random Forest trained with k-fold validation

80:20 split

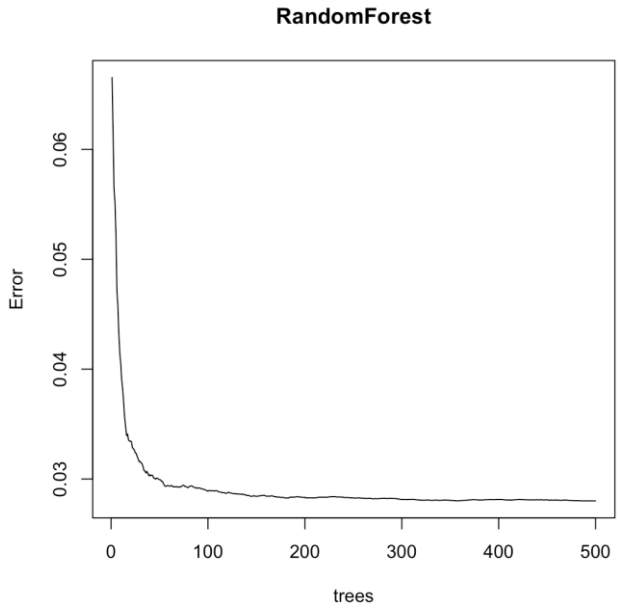


Fig. 79: error distribution as a function of the trees

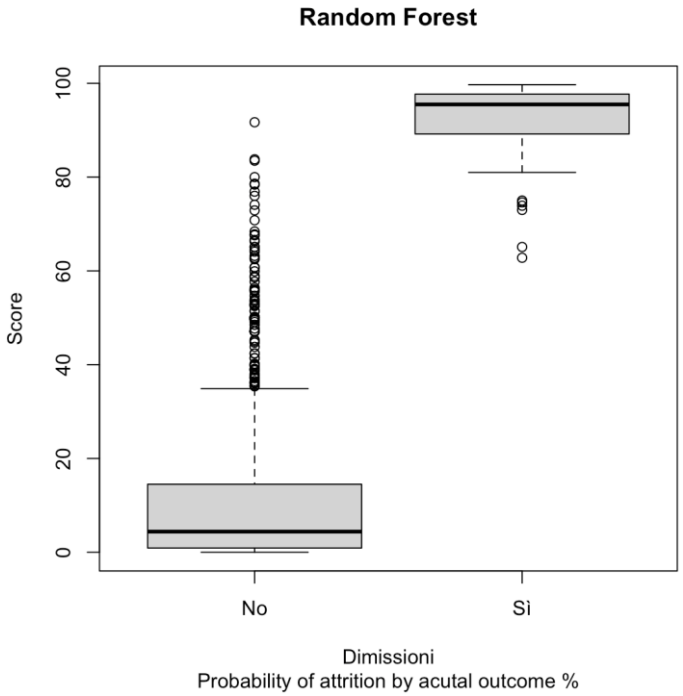


Fig. 80: distribution of the Score on the test set

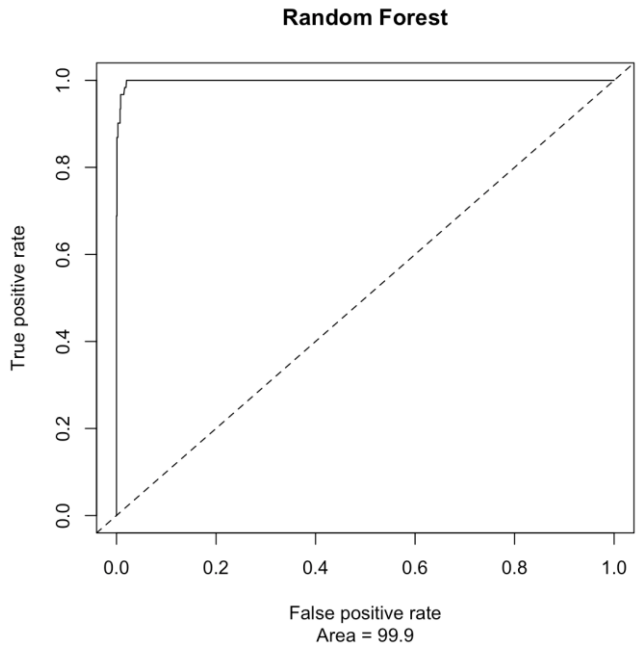


Fig. 81: Random Forest's ROC curve

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  1030    0
     TRUE     61   61

               Accuracy : 0.947
                 95% CI : (0.9325, 0.9593)
    No Information Rate : 0.947
    P-Value [Acc > NIR] : 0.534

                  Kappa : 0.6413

 Mcnemar's Test P-Value : 1.564e-14

            Sensitivity : 0.9441
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 0.5000
             Prevalence : 0.9470
         Detection Rate : 0.8941
   Detection Prevalence : 0.8941
      Balanced Accuracy : 0.9720

       'Positive' Class : FALSE
```

Fig. 82: Random Forest's Confusion Matrix and Statistics

```
Random Forest

4615 samples
  14 predictor
   2 classes: 'No', 'Sì'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 4154, 4154, 4153, 4153, 4153, 4153, ...
Resampling results across tuning parameters:

  mtry  ROC        Sens       Spec
   2    0.9130985  0.7994286  0.8756171
  25    0.9971997  0.9526827  0.9815921
  48    0.9954314  0.9413734  0.9847311

ROC was used to select the optimal model using the largest value.
The final value used for the model was mtry = 25.
```

Fig. 83: Performance measures of the Random Forest trained with k-fold validation

Variable importance

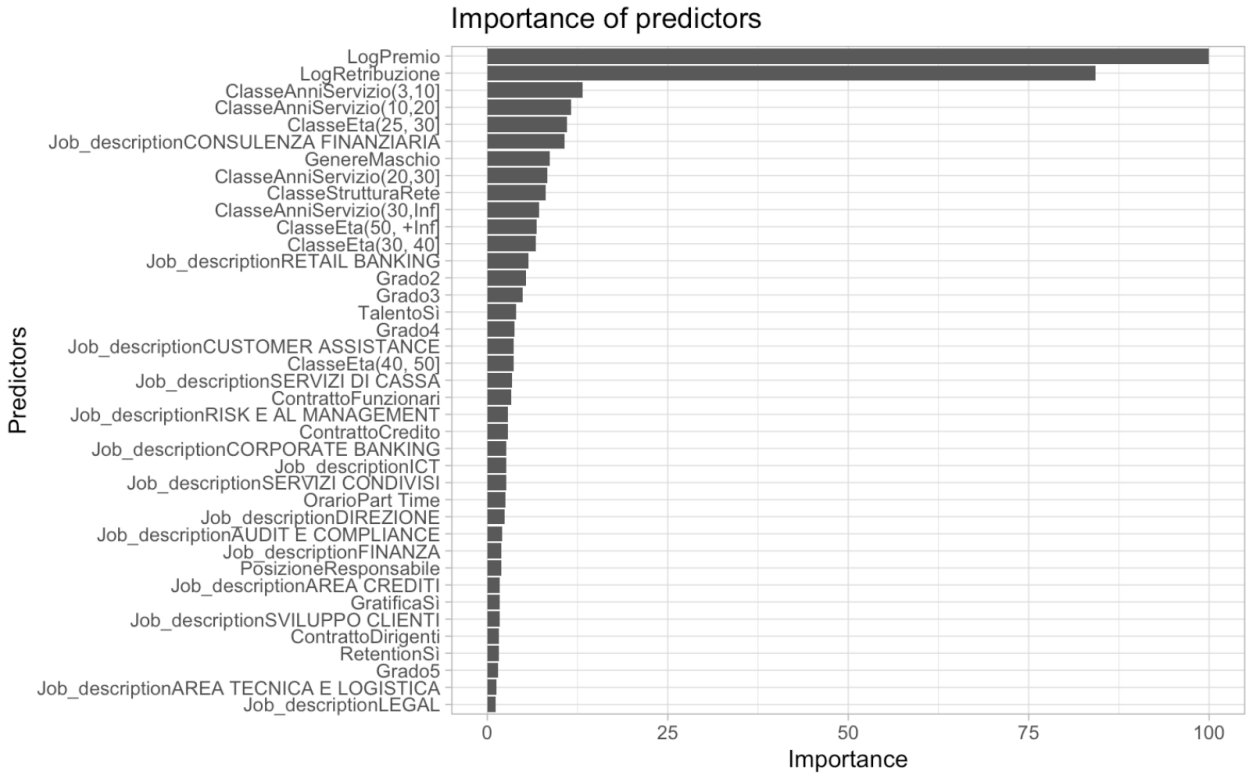


Fig. 84: Random Forest's variable importance

4.3 Model Evaluation and discussion

This phase evaluates the qualities of the adopted models. In this section, we provide a comparison of all models on the 80/20 and 70/30 splits. Table 86 shows the performance results of the predictive algorithms that we have implemented.

Model Type		Split on Train and Test set	Values of Criteria (%)			
			AUC on Test	Sensitivity on Test	Specificity on Test	Accuracy on Test
Logistic Regression	LASSO	80:20	85,2	72,1	77,3	77
	RIDGE		85	83,6	68	68,8
	MIX		85,4	72,1	74,8	74,7
	LASSO	70:30	86,4	77,2	79,7	79,6
	RIDGE		86,2	79,3	77	77,1
	MIX		86,6	77,1	79	78,9
Naïve Bayes		80:20	81,6	73,8	74,9	74,8
		70:30	83,5	80,4	75,4	75,6
Decision Tree		80:20	86,8	76,1	82	76,3
		70:30	87,2	82	76,1	81,7
Random Forest		80:20	99,9	94,4	100	94,7
		70:30	99,7	95,8	97,8	95,9

Fig. 85: performance metrics of the adopted models

According to our results, the best performing algorithm is the Random Forest with a 80:20 split between training and test set; for this reason we focused on the results of this algorithm to identify the main groups of employees who have a high risk of attrition, which are those on which it would be useful to identify interventions to reduce voluntary resignation.
To understand which variables have the greatest predictive power, we used the varImp function of the Caret package (Kuhn, 2008). This method combines two measures to assess the importance of each factor (Hoare, 2018). The first is simply how much accuracy decreases when a variable is excluded. The second is based on the gain of the target function (*i.e.* the Gini index) when choosing a variable to split a node; this is a measure of the probability of incorrectly classifying a new instance of a random variable.

Our analyses show that Random Forest, with an 80:20 split between training and test set, produced the best results for the case study. We therefore assessed the importance of each factor in predicting employee attrition with the Random Forest algorithm and we found that the results are in line with those of the exploratory analyses, and gave us other important information. As can be seen in Fig. 8, the results show that the monetary incentive is very important: both the premium ("Log Premium") and the remuneration ("Log Salary") are among the most important predictors of employee attrition. Since the premium is a remuneration for activities that go beyond those that are contractually defined, it is a component that can also denote a particular commitment by the employee. The subsequent most relevant variable is the job description, and in the following part of this chapter we will move on to describe in greater detail the relationship between the various positions and the probability of resigning. The four most relevant component is seniority ("Tenure

class"); this shows that people who have been at the company for a long time tend to stay. The age ("Age class") is also decisive, as younger employees are more likely to leave the organization. Next, another significant variable is the contractual level ("Level"). In addition, gender plays an important role, male workers are more likely to resign, and exploratory analyses allow us to understand that it is related to the fact that they aspire more to positions they perceive as better. Next, the most important components are the type of contract, the department and the rank. Finally, being part of a talent program is also associated with a greater propensity to resign, probably because they know they have value on the labour market.

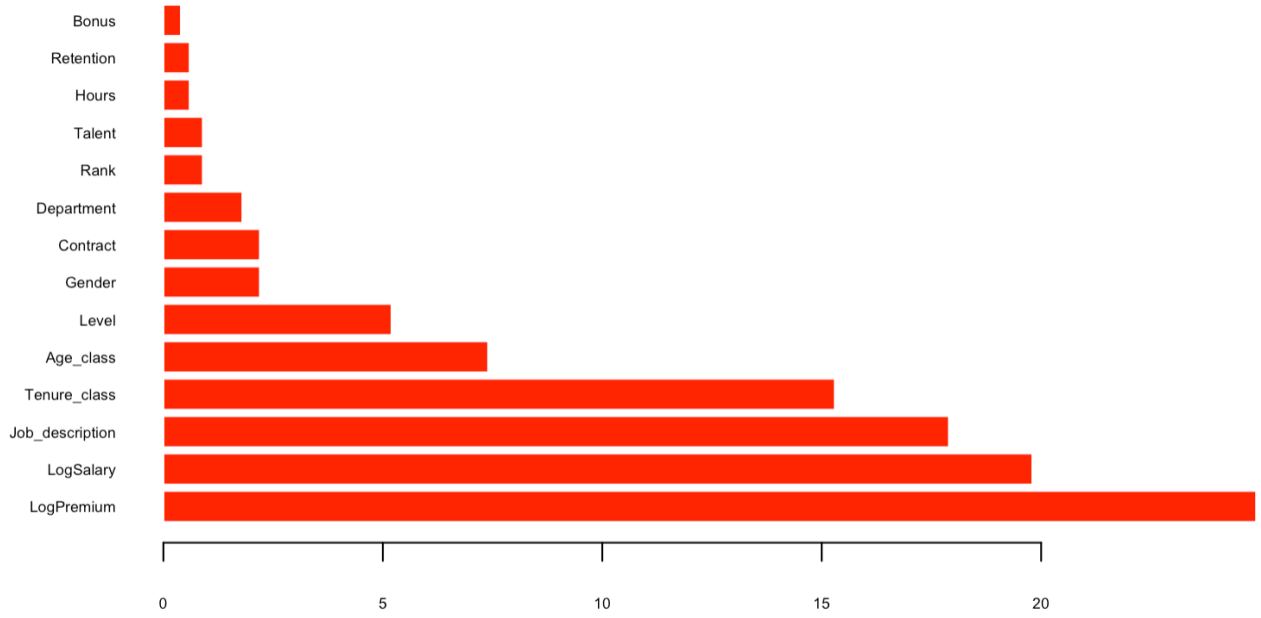


Fig. 86: Predictors by relative importance (%)

Once the overview with the variable importance were understood, we computed the SHAP method to understand specifically which areas would be more advantageous in terms of costs and benefits to intervene. It was performed using the R iml package, an extended version of the Kernel SHAP method for approximating Shapley values. It is of interest when attempting to explain complex machine learning models, as the random forest in our case. The SHAP (SHapley Additive exPlanations) algorithm was first published in 2017 by Lundberg and Lee, and is a unified framework for interpreting predictions (Lundberg and Lee, 2017; Mazzanti, 2020). To accomplish this, it quantifies the contribution that each feature brings to the prediction made by the model (see Appendix for more information about the SHAP method).

Using the R iml package, we plotted the graph. On the x-axis it shows the Shap values: the values on the left represent the observations that shift the predicted value in the negative direction, while the values on the right contribute to shift the predicted value in a positive direction. On the left y-axis are represented all the features (Choudhary, 2019).

The following graph shows the relationship between the most important variable, the logarithm of the premium, and the probability of leaving the company. It shows that the relationship between the two variables is inverse: as the logarithm of the bonus increases, the probability of leaving decreases. The same applies also to the logarithm of the salary, although to a more modest extent.

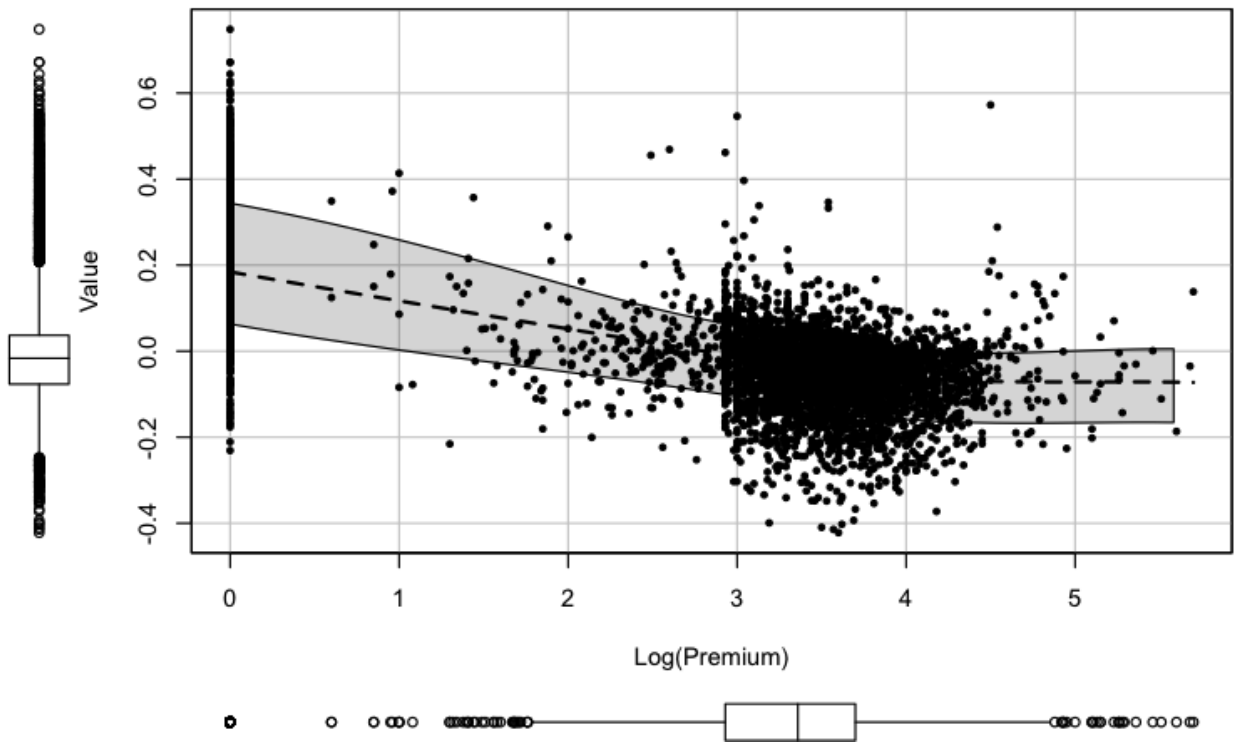


Fig. 87: Relationship between the logarithm of the premium and the probability of attrition

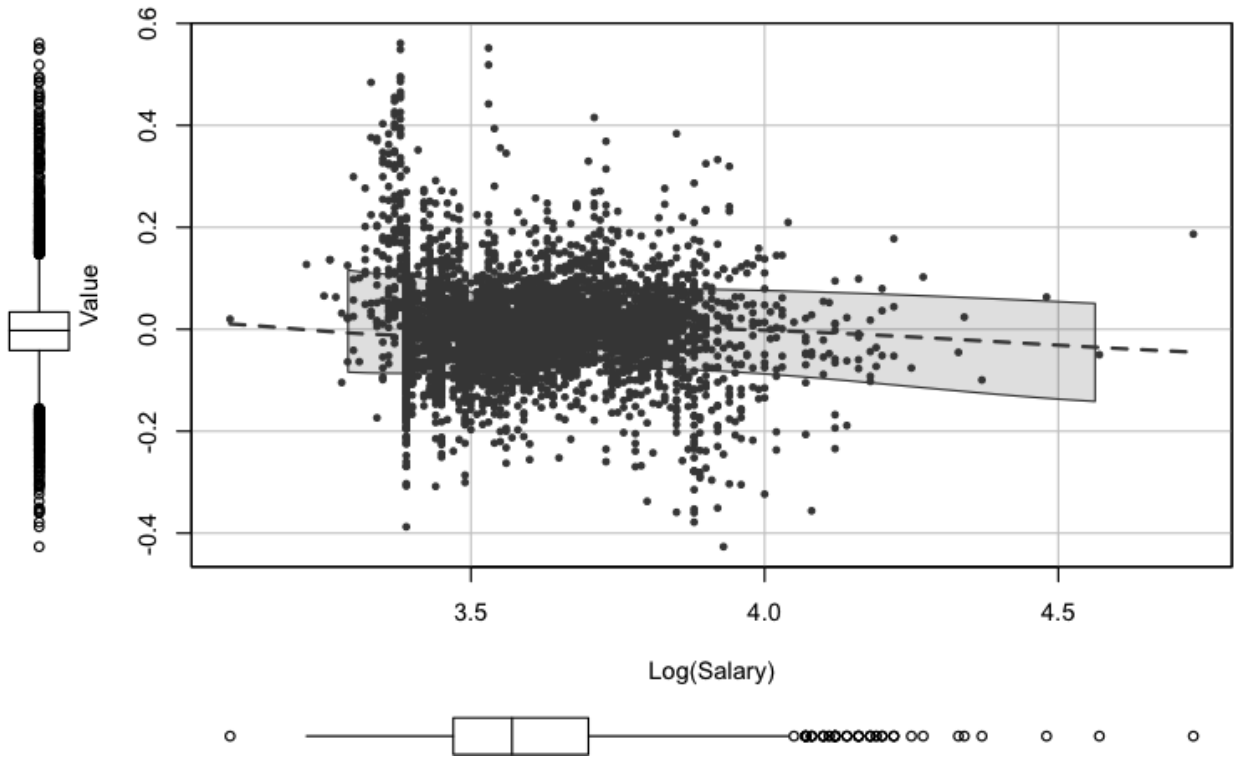


Fig. 88: Relationship between the logarithm of the salary and the probability of attrition

The following graph shows the trend of attrition according to the third most important variable that is the tenure, evaluated according to the five classes we had defined. The results show that, for classes up to 10 years, the contribution to the probability of attrition is positive, while those ranging from 10 years onwards have as a central trend a zero or negative contribution.

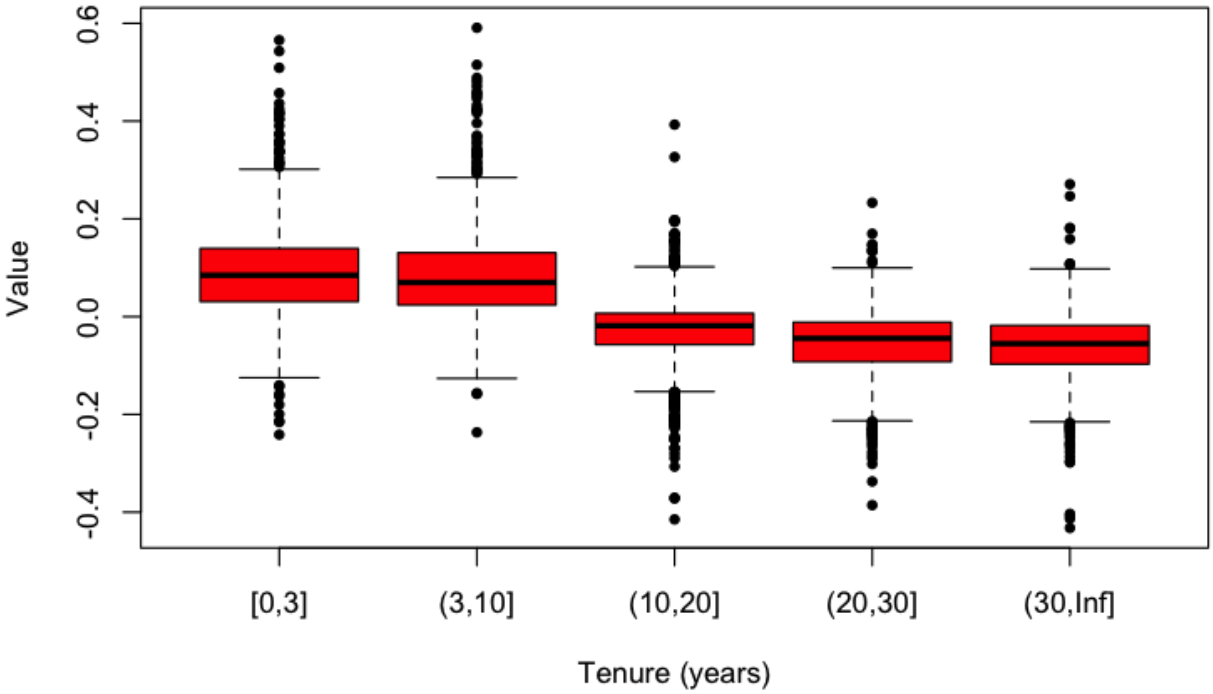


Fig. 89: Relationship between the logarithm of the tenure and the probability of attrition

Another very important variable is the job description. For reasons of data confidentiality, we have used numbers to indicate job descriptions. We focus on some significant areas to study their challenges.

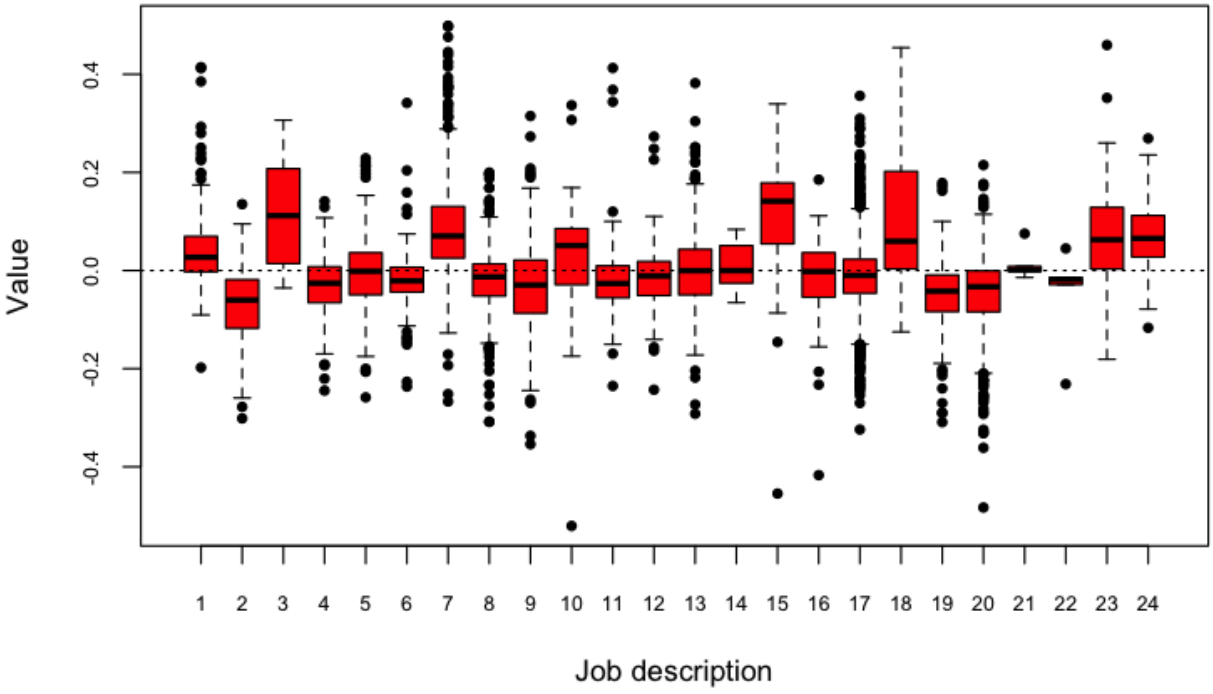


Fig. 90: Relationship between the job description and the probability of attrition

Area 24 was chosen as an example, which is a case in which the Shap values are relatively positive and includes few employees. The following tables show that the probability of leaving the company in this area is medium for the youngest, high between 3 and 10 years, and low after 10 years of company seniority. These findings indicate that, for this position, the competitive advantage on the labour market is acquired after a few years of experience. In addition, those who are employed under a contract in the credit area have a high probability of leaving the company.

Table 4: Attrition rates by tenure class

Tenure class	Probability of attrition (%)
[0, 3]	69.7
(3, 10]	99.5
(10, 20]	12.7

Table 5: Attrition rates by contact

Contract	Probability of attrition (%)
Other	41
Credit	99.4

Managers	11.1

The graph below shows the trend in the probability of attrition based on the logarithm in base 10 of the total salary of the employees (given by the sum of the premium and the basic salary). We can notice that those who have the lowest income have a fairly high probability of leaving, and as Log income grows the probability drops significantly, unless a few cases that deviate.

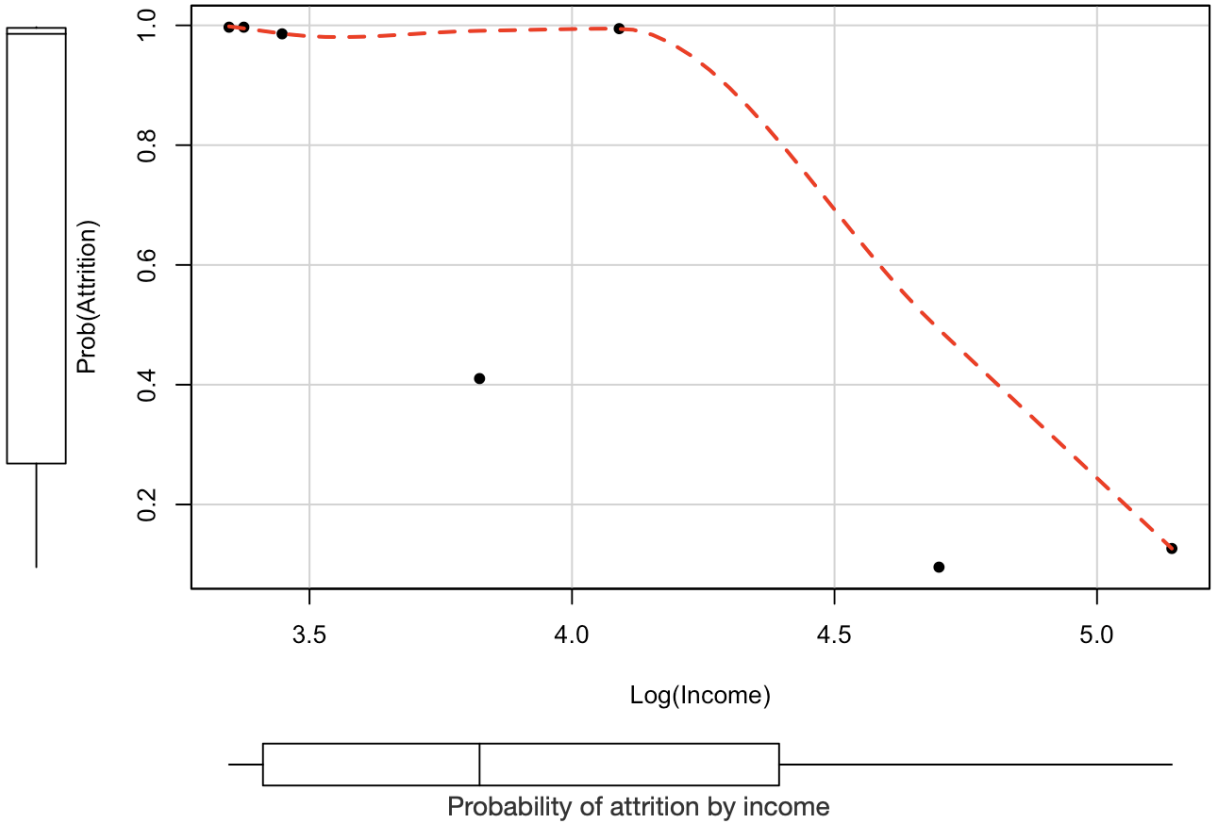


Fig. 91: Relationship between the logarithm of income and the probability of attrition

As an example, we take the case of a single employee, to show how the results are able to suggest what interventions would be convenient to do to retain him. The employee in question is relatively old (between 30 and 40 years old), has low tenure, has no retention incentive, and has average total compensation.
By calculating Shap values, we can see which factors contribute most significantly to the probability that the employee resigns (which in this case is high, close to the unit).

Table 6: Detail of the feature value

Feature value
Contract = Other
LogSalary = 3.82
Position = accountable
Talent = No

Retention = No
Hours = Full Time
Bonus = No
Department = Center
Level = 2
Gender = Male
Tenure class = [0,3]
Age class = (30,40]
LogPremium = 0
Job description = 24

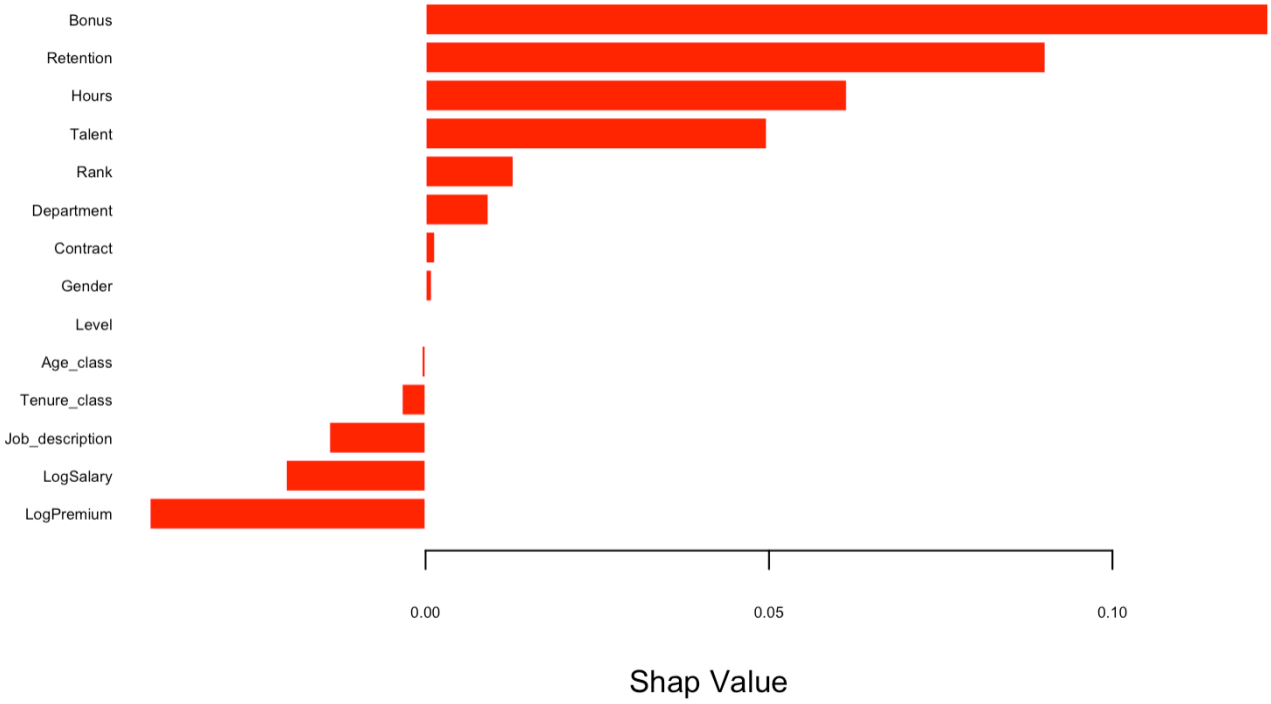


Fig. 92: Shap values of the example

Table 6 and Figure 92 show the marginal contributions of the different variables, and how much they affect the probability of leaving. These values give an indication of the areas in which action can be taken to reduce the likelihood of the employee leaving the company. As the values of Premium and Salary increase, the probability of the employee leaving decreases marginally. Salary helps retain employees and this is a lever on which the organization can act.

As the values of Bonus, Retention, Hours and Talent increase, the probability of the employee leaving the company increases.

As for retention, looking at this graph it would seem that giving a retention incentive increases the probability of leaving. The positive relationship between Retention and Attrition is the opposite of what might appear and is explained by the fact that Retention is an economic incentive that is given by the organization to those workers it believes are likely to leave and that it doesn't want to let go.

The fact that attributing a retention incentive has a positive effect on attrition means that the organization is actually identifying a person who is likely to leave, and gave a retention incentive to those people who were more likely to leave.
The data referring to the fact that a worker has received a retention incentive is static: it may happen that a worker has received this incentive as soon as he/she entered the employment relationship with the company, which remains in the database even when it expires. These retention incentives last for a certain period of time, and when they cease, the person is likely to leave. This justifies the fact that those with a retention incentive may leave the company at some point. This means that the organization's policy of giving retention incentives works. The fact that the probability of leaving is positively associated, other things being equal, with the fact that an employee has been given a retention incentive means that the organization has actually understood to which employees to assign it.
The premium is assigned to those workers who carry out work activities that go beyond their contractual obligations. The inverse causality is explained by the fact that taking a premium is an indicator of the fact that the employee has a strong commitment to his or her employer, economically recognised.
With regards to the Hours variable, the graph shows that those who have a full-time commitment are more likely to leave than those who have a part-time. An explanation of the causal relation could be that there are some people who would like to have a lower commitment in terms of working hours and are not able to get it within the organisation, so they look for it outside. As can be seen from table 8, part-time commitment is preferred by female workers, even if it is always in the minority compared to full-time.

Table 7: Gender rates by full-time/part-time commitment (%)

	Full-time	**Part-time**
Female	86.0	14.0
Male	99.4	0.6

From the tables above we can see that the type of part-time contract is prevalent in the lowest career brackets, which are more frequently covered by women. It seems that this type of worker, who does not see great opportunities for career development, wishes to have more time to devote to private life and aspires to part-time working. If the organisation does not grant them this, they look for this type of position elsewhere. To ensure that these workers do not leave, the organisation could reduce their working hours.

Table 8: Gender rates by full-time/part-time commitment, type of contract = "Other" (%)

	Full-time	**Part-time**
Female	2.5	0.1
Male	8.3	0.1

Table 9: Gender rates by full-time/part-time commitment, type of contract = "Credit" (%)

	Full-time	Part-time
Female	69.9	13.3
Male	62.0	0.3

Table 10: Gender rates by full-time/part-time commitment, type of contract = "Managers" (%)

	Full-time	Part-time
Female	0.7	0.0
Male	5.7	0.1

Table 11: Gender rates by full-time/part-time commitment, type of contract = "Executives" (%)

	Full-time	Part-time
Female	12.9	0.5
Male	23.4	0.1

Talent is a recognition that is given to employees based on what is the potential of development of their skills and productivity within the company.
The relationship between the positive marginal impact of talent and the probability of leaving could be explained by the fact that actually the organization is able to identify resources that have a high potential and these here are the ones that are more likely to look for job opportunities outside the company.
Indeed, incorporating certain people in a privileged career programme for talent is not enough to retain them completely. It would seem, looking at these results, that those with a talented qualification, while knowing that they may have a privileged prospect of career advancement in the organisation, are also inclined to look for other opportunities outside. It seems that this is due to the fact that those who belong to the talent programme do not have a particularly higher economic recognition than those who do not (Fig. 14). Thanks to discussions with the organisation, we have been able to confirm that the talent programme does not bring any financial incentives per se, it is just a potentially faster career path. In addition, most of the people in the talent programme are young, so they are at an age when they feel they have a chance of finding opportunities on the labour market. Moreover, they have a rather low tenure, and another explanation for their propensity to leave the organisation may be that they have not invested as many years in the organisation and perceive the cost associated with leaving it to be lower.

Table 12: Talent rates by tenure (%)

	[0, 3]	(3, 10]	(10, 20]	(20, 30]	(30, Inf]
No	16.1	14.7	37.5	20.0	11.7
Yes	27.9	40.7	24.6	4.7	2.0

Table 13: Talent rates by age (%)

	[0, 25]	(25, 30]	(30, 40]	(40, 50]	(50, Inf]
No	3.1	20.5	26.0	30.1	20.3
Yes	4.4	63.3	22.9	4.7	4.7

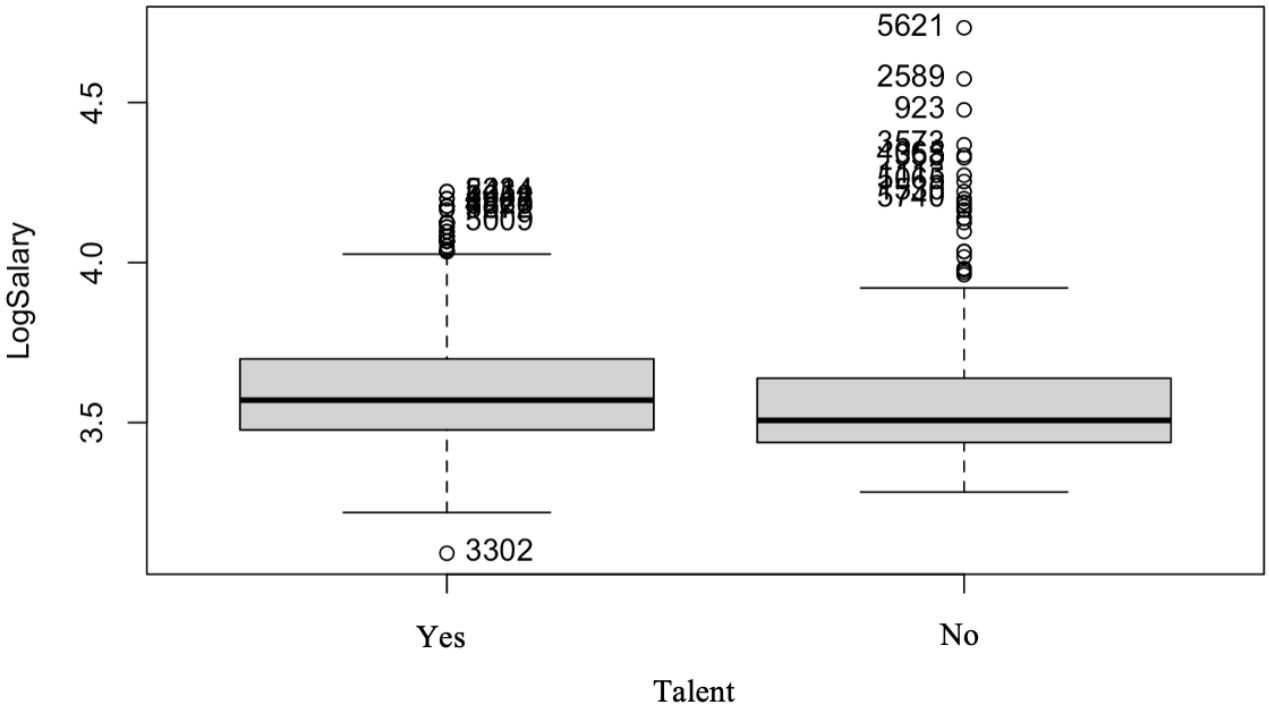


Fig. 93: Logarithm of the Salary given Talent (%)

As we have seen from the interpretation of the mechanisms of casuality, they may not be obvious and are interpretable only with respect to the specific context from which these data come. In fact, if we interpret the values of the graph without knowing how the source data are structured and what they indicate, we could get to misinterpret the relationship of causality and then we could make the wrong conclusions that could lead to making counter-productive interventions.

Another significant example is the second job description.
The following tables show that the tenure is inversely proportional to the probability of leaving, and younger employees and those belonging to the credit contract area are more likely to leave.

Table 14: Attrition rates by tenure class

Tenure class	Probability of attrition (%)
[0, 3]	46.5
(3, 10]	14.2
(10, 20]	7.5
(20, 30]	0.7

(30, +Inf]	1

Table 15: Attrition rates by age class

Age Class	Probability of attrition (%)
(25, 30]	40.5
(30, 40]	20.6
(40, 50]	3.6
(50, +Inf]	3.7

Table 16: Attrition rates by contact

Contract	Probability of attrition (%)
Other	3.6
Credit	28.4
Managers	1.7
Executives	10.2

The graph below shows the trend in the probability of attrition based on the logarithm in base 10 of the total salary of the employees (given by the sum of the premium and the basic salary). We can notice that the probability of leaving is much lower than in the area we previously analyzed and decreases as the logarithm of income increases. Income decreases more rapidly for lower values of the logarithm, and as the logarithm increases, the rapidity of decrease declines. Most cases belong to the area that is below 20% of the value of the probability of attrition. We focus on the 13 cases that have an attrition probability greater than 0.5, as they are those at highest risk of leaving.

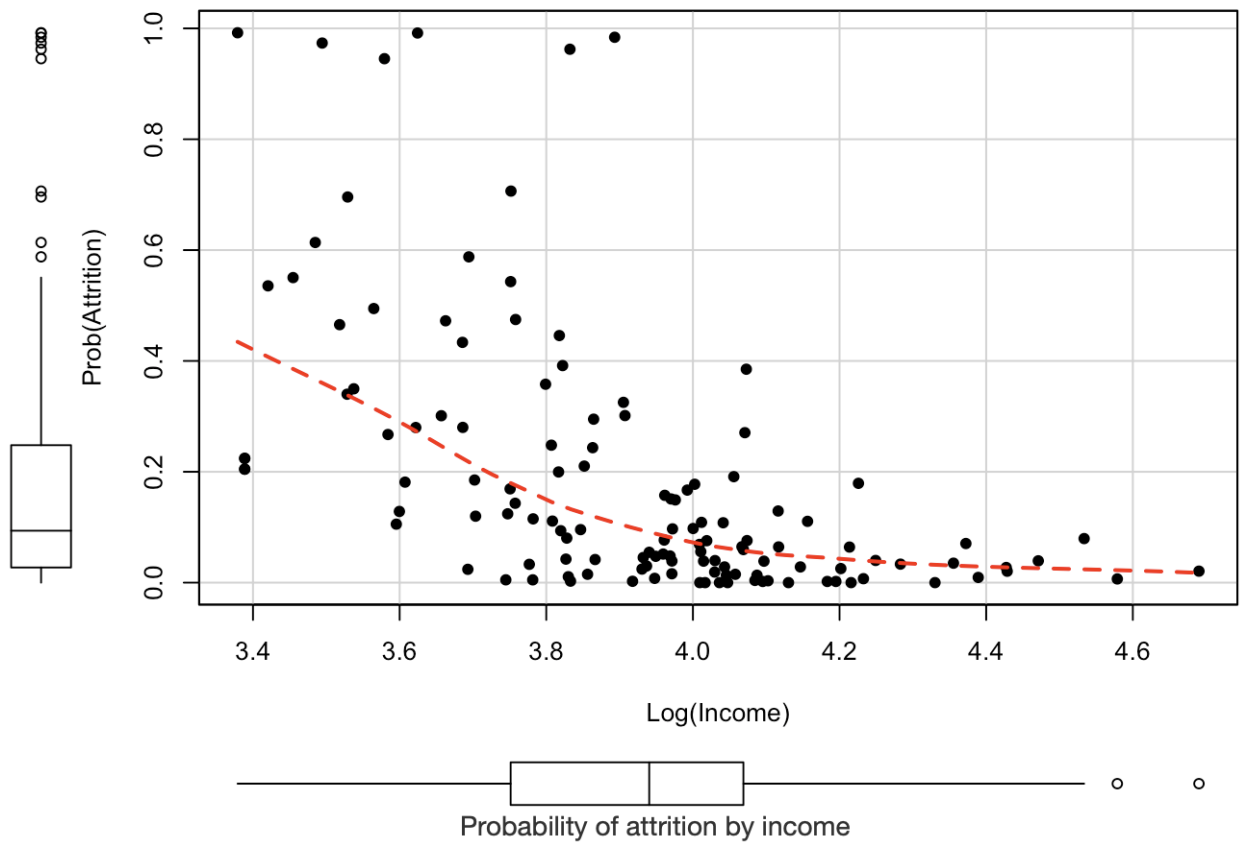


Fig. 94: Relationship between the logarithm of income and the probability of attrition

Since in this case the number of marginal effects is 182 (13 observations x 14 variables), it is not possible to represent them with a table and a graph as in the previous case.
We then select some features and find out what are the average marginal effects on the probability of quitting. Table 18 and Fig. 17 show that the values of the features that contribute most to increasing the probability of leaving in this area are having a company seniority between 0 and 3 years, which increases by 18% the probability of leaving, not receiving a premium, for the same reason we saw in the previous area, and being a male worker.

Table 17: average marginal effects of the feature value

Feature value	**Marginal effect (%)**
Gender = Male	2.9
Tenure class = [0,3]	18
Contract = Credit	0.04
LogSalary = 3.48	0.2
Position = responsible	0.6
Talent = No	0.1
Retention = No	0

Hours = Full Time	0.3
Bonus = No	0.06
Department = Center	2.6
Level = 2	1.7
LogPremium = 0	25

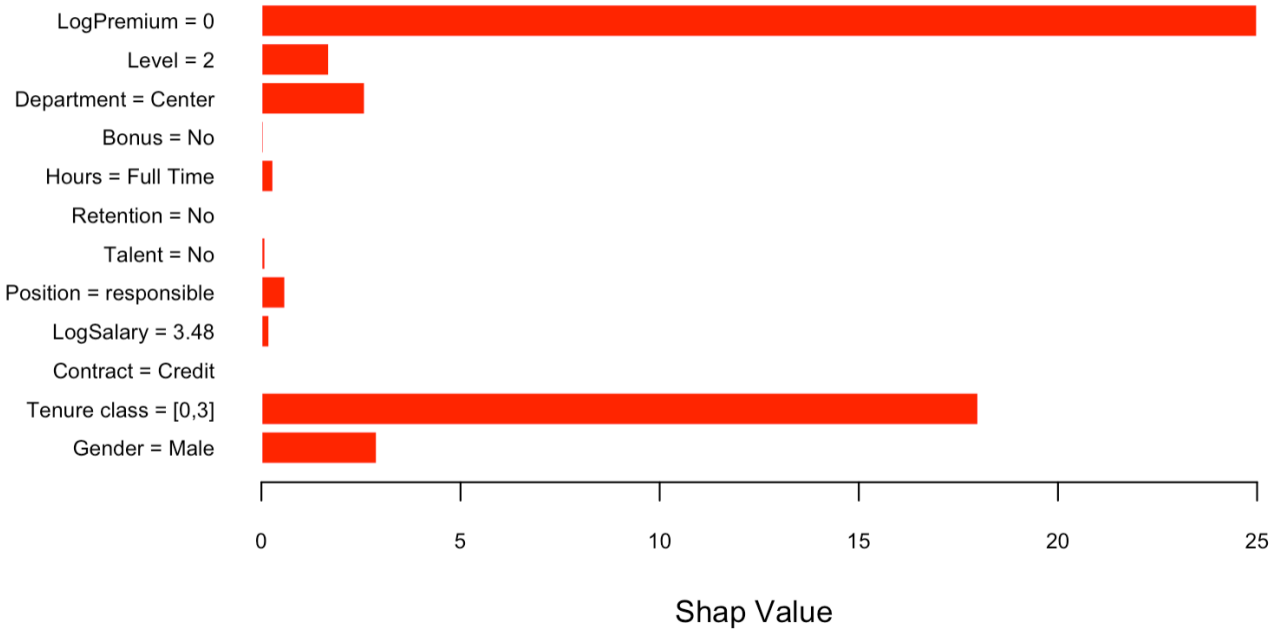


Fig. 95: Shap values of the examples

5. Conclusions

This study investigates employee attrition on a real dataset from an Italian financial company. We reviewed the literature about predicting employee attrition and its motivations and identified what are the main predictors and motivations of employee attrition. We also provided an overview of the most common and promising machine learning models that appear in the extant research. For the purpose of this study, we selected four models, namely: Naïve Bayes, Logistic Regression with Elasticnet penalty, Decision Tree and Random Forest. We compared the results of these models according to standard metrics and selected Random Forest as the best one for our dataset. Random Forest trained on 80% of our data and tested on the remaining 20% achieves 99.9% of Area under the ROC graph and 94.4% of Sensitivity.

The dataset on which we performed our analysis is strongly unbalanced; indeed, only 5.4% of examples are positive, that is resigned over the sampling period. Therefore, we combined Random Forest with ROSE to overcome the class imbalance problem.

Our results show that machine learning techniques together with careful data analysis can be used to build reliable and accurate predictive employee attrition predictions.

We also provided sample insights about what are the main features of some employees leaving the company, based on SHAP values analysis. These findings are of interest to the human resource department of the company which, thanks to the predictions and feature analysis of employees leaving the company may establish retention plans to mitigate the risk of attrition.

A limitation of this study is that resignation and stay cases are not observed over the same period of time. This is because the data that were actually made available to our analysis from the company are not synchronous because of its privacy policy. Therefore, we had to make the assumption that the probability of attrition is time independent. Also, some constructs that are deemed important according to the extant literature could not be considered in our predictive models because of lack of data.

Our future research agenda points to improving the dataset and considering other machine learning algorithms which may improve upon the current results. In this sense, candidate models are XGBoost and Neural Networks.

Appendix: the SHAP method

There are some models that are easy to understand and interpret in terms of their importance and partial effects of features on the predicted outcome; examples are generalized linear models, such as logistic regression. While not necessarily essential, interpretability is often useful to evaluate what actions an organization can take to improve its processes and activities. However, there is sometimes a trade-off between interpretability and predictive performance; for example, in our case random forest, where the roles of any features are not easy to interpret, ranks above logistic regression, where they are.

Various methods have been proposed to interpret the prediction of complex models, that consist of an interpretable approximation of the original model. Lunberg and Lee (2017) show that the SHAP Value method "is better aligned with human intuition as measured by user studies and more effectually discriminate among model output classes than several existing methods". SHAP Value assigns each feature an importance score for each prediction based on the notion put forward by Shapley in the context of cooperative game theory (Shapley, 1951; Kaur *et al.*, 2020; Choudhary, 2019). SHAP Value quantifies the marginal contribution that each feature brings to the prediction of the model; the "players" are the features and the "game" is evaluating the target variable (Mazzanti, 2020). The SHAP method simulates that only some features are present (playing) and some are absent (not playing); therefore the differential contribution of each feature is evaluated.

For example, Mazzanti (2020) illustrates a machine learning model that predicts the income of a person knowing the age, gender and job. To determine the importance of a single player, each possible combination of players should be considered, meaning each possible combination of features, their "power set". The power set can be represented as a tree, where each node is a coalition of features, and each edge points to a node where a feature not present in the previous coalition is added (Fig. 18). A predictive model for each distinct coalition in the power set is trained and the differential contribution of each additional feature to the outcome is evaluated. Then, the marginal contribution of each feature is measured as the average of the effects on the target

variable across all instances over the tree where it is added. For example, the contribution of age to income is the average effect of adding nodes 2 to 1, 5 to 3, 6 to 4, 8 to 7.

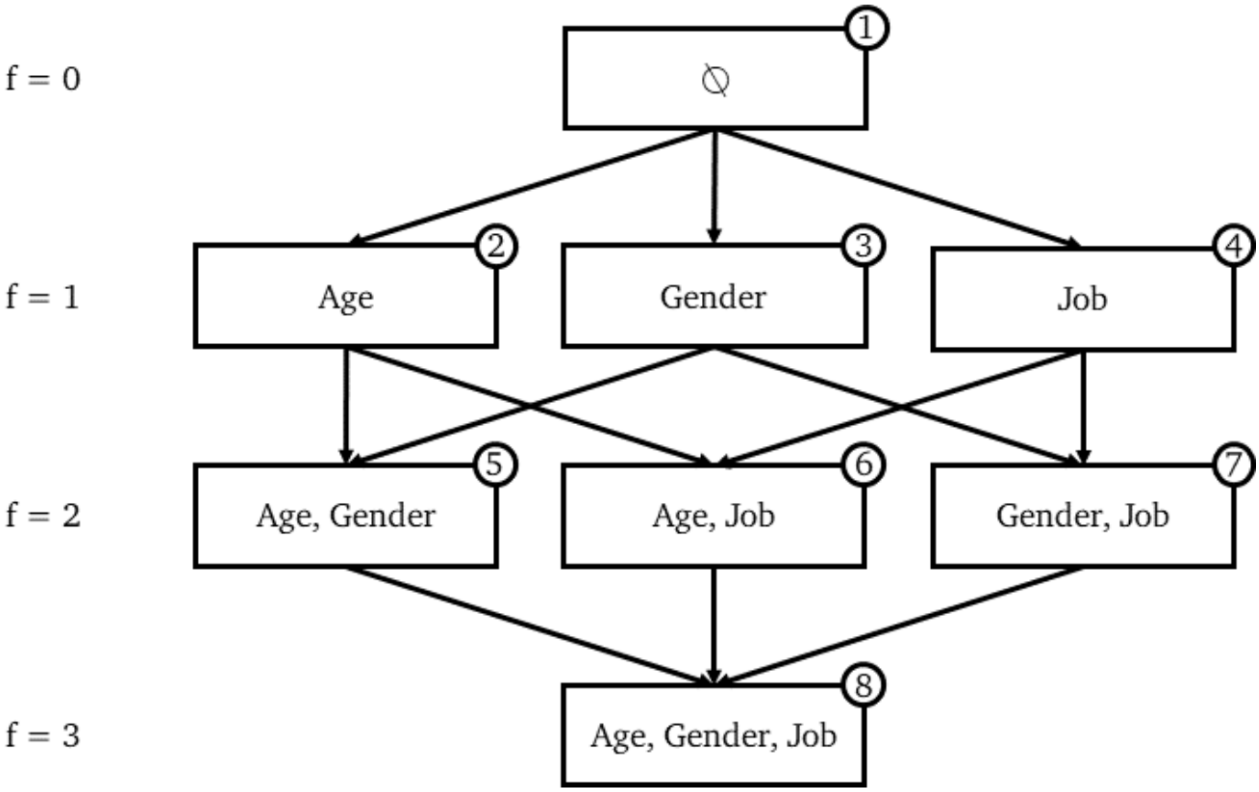


Fig. 96: Power set of features (Source: Mazzanti, 2020)

References

Abeble, G. (2016). Causes and Consequences of Employee Turnover. The Case of Kolfie Keranio Sub-City. [Master dissertation, St. Mary's University]. *St. Mary's University Institutional Repository.* http://hdl.handle.net/123456789/3780

Akinyomi, O. J. (2016). Labour Turnover: Causes, Consequences and Prevention. *Fountain University Journal of Management and Social Sciences (Special Edition), 5*(1), 105-112.

Alao, D. A. B. A., & A. B. Adeyemo (2013). "Analyzing employee attrition using decision tree algorithms." Computing, Information Systems, Development Informatics and Allied Research Journal 4.1: 17-28.

Alduayj, S. S., & Rajpoot, K. (2018). Predicting Employee Attrition using Machine Learning. *2018 International Conference on Innovations in Information Technology (IIT)*, 93–98. https://doi.org/10.1109/INNOVATIONS.2018.8605976

AlSayyed, B., & Braiki, F.A. (2015, March 3-5). Employee turnover, causes, the relationship between turnover and productivity and recommendations to reduce it. *Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management Dubai, United Arab Emirates (UAE)*.

Arokiasamy, A. R. A. (2013). A Qualitative Study on Causes and Effects of Employee Turnover in the Private Sector in Malaysia. *Middle-East Journal of Scientific Research 16*(11), 1532-1541. https://doi.org/10.5829/idosi.mejsr.2013.16.11.12044

Ayuure, A. A. (2013). Causes and Effects of Employee Turnover in the Commission on Human Rights and Administrative Justice in the Upper East Region of Ghana. [Master dissertation, University of Cape Coast]. *University of Cape Coast Institutional Repository.*

Balcha, S. M. (2019). Causes and Consequences of Employee Turnover in International Medical Corps Ethiopia Mission. [Master dissertation, St. Mary's University]. *St. Mary's University Institutional Repository.*

Bennett, N., Blum, T. C., Long, R. G., & Roman, P. M. (1993). A Firm-Level Analysis of Employee Attrition. *Group & Organization Management*, *18*(4), 482–499. https://doi.org/10.1177/1059601193184006

Chang, H. Y. (2009). Employee turnover: A novel prediction solution with effective feature selection. WSEAS Transactions on Information Science and Applications, 6(3), 417-426. ISSN: 1790-0832.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, C., Liaw, A. & Breiman, L. (2004). *Using Random Forest to Learn Imbalanced Data*. Report Number 66. Berkeley Statistics, Research and Industry.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Expert Systems with Applications, 34, 280–290. https://doi.org/10.1016/j.eswa.2006.09.003

Choudhary, A. (2019). A Unique Method for Machine Learning Interpretability: Game Theory & Shapley Values. Analytics Vidhya. https://bit.ly/3Dweoqz

Cohen, A. (1993). Age and Tenure in Relation to Organizational Commitment: A Meta-Analysis. *Basic and Applied Social Psychology*, *14*(2), 143–159. https://doi.org/10.1207/s15324834basp1402_2

Colding, Linda K. (2006). Will They Stay or will They Go? Predictors of Academic Librarian Turnover. In *Advances in Library Administration and Organization, 23,* 263–280. https://doi.org/10.1016/S0732-0671(05)23007-9

El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, *28*(6), 1273–1291. https://doi.org/10.1108/IJOA-10-2019-1903

Esmaieeli Sikaroudi, A., RouzbehGhousi, & Sikaroudi, A. (2015). A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing). *Journal of Industrial and Systems Engineering*, *8*.

Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting Employee Attrition Using Machine Learning Techniques. *Computers*, *9*(4). https://doi.org/10.3390/computers9040086

Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting Employee Attrition Using Machine Learning Techniques. *Computers*, *9*(4), 86. https://doi.org/10.3390/computers9040086

Fareri, S., Chiarello, F. Coli, E., Fantoni, G., Binda, A., (2020). Estimating industry 4.0 impact on job profiles and skills using text mining. Computers in Industry, Volume 118, 103222, ISSN 0166-3615. https://doi.org/10.1016/j.compind.2020.103222

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Getachew, T. (2016). Assessment of Employee Turnover Causes at Ethiopian Revenues and Customs Authority [Master dissertation, St. Mary's University]. *St. Mary's University Institutional Repository.* http://hdl.handle.net/123456789/3912

Getachew, Y. (2017). The Causes and Impact of Employee Turnover on Project Performance. The Case of Ethiopia Sugar Corporation Projects. [Master dissertation, St. Mary's University]. *St. Mary's University Institutional Repository.* http://hdl.handle.net/123456789/3224

Hoare, J. (2018). How is Variable Importance Calculated for a Random Forest? https://www.displayr.com/how-is-variable-importance-calculated-for-a-random-forest/

Hochwarter, Wayne A., Perrewé, P. L., Ferris, G. R., & Guercio, R. (1999). Commitment as an Antidote to the Tension and Turnover Consequences of Organizational Politics. *Journal of Vocational Behavior*, *55*(3), 277–297. https://doi.org/10.1006/jvbe.1999.1684

HR Analytics- exploration and modelling with R. https://www.kaggle.com/ragulram/hr-analytics-exploration-and-modelling-with-r

Hussein, A. H. Al-A. (1989). Labour turnover in the West Bank: an analysis of causes of turnover in the industrial sector. [Doctoral dissertation, University of Glasgow]. *Glasgow Theses Service.* http://theses.gla.ac.uk/id/eprint/2850

Imani, Z. (2013). Assessment of the Causes of Labour Turnover in Organizations in Tanzania: a Case of National Bank of Commerce (NBC). [Master dissertation, Mzumbe University]. *Unimib Digital Repository.* http://hdl.handle.net/11192/1950

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. doi:10.1145/3313831.3376219

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1 - 26. doi:http://dx.doi.org/10.18637/jss.v028.i05

Lantz, B. (2019). Machine Learning with R. Expert techniques for predictive modeling. Third Edition. *Packt Publishing Ltd.*, ISBN 978-1-78829-586-4.

Lazzerini, B. (2021). Intelligent Systems Master Course (University of Pisa) [PowerPoint slides]. Retrieved from https://elearn.ing.unipi.it/course/index.php

Liaw A. and Wiener M. (2002). Classification and Regression by randomForest. R News, 3(3), 18-22. https://CRAN.R-project.org/doc/Rnews/

Lunardon, N., Menardi G., & Torelli, N. (2014). ROSE: a Package for Binary Imbalanced Learning. The R Journal, 6(1). ISSN 2073-4859.

Lundberg, S. M, and Lee S. I. (2017). A Unified Approach to Interpreting Model Predictions. arXiv:1705.07874

Majka M (2019). naivebayes: High Performance Implementation of the Naive Bayes Algorithm in R. R package version 0.9.7, https://CRAN.R-project.org/package=naivebayes

Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, *128*(584), 2145–2166. https://doi.org/10.1256/003590002320603584

Mazzanti, S. (2020). Shap Values Explained Exactly. How you Wished Someone Explained to You. Towards data science https://bit.ly/3nbPpUH

McQuerrey, L. (2019). Employee Turnover Vs. Attrition. https://smallbusiness.chron.com/employee-turnover-vs-attrition-15846.html

Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, *28*(1), 92–122. https://doi.org/10.1007/s10618-012-0295-5

Milborrow, S. (2020) Plotting rpart trees with the rpart.plot package. http://www.milbo.org/rpart-plot/prp.pdf

Molnar, Christoph, Giuseppe Casalicchio, and Bernd Bischl. "iml: An R package for interpretable machine learning." Journal of Open Source Software 3.26 (2018): 786.

Nappinnai, M. V., & Premavathy, N. (2013). Employee Attrition and Retention in A Global Competitive Scenario. *International Journal of Research in Business Management (IMPACT: IJRBM) ISSN(E): 2321-886X; ISSN(P): 2347-4572, 1*(6), 11-14.

Negassa, S. B. (2016). Antecedents and Consequences of Employee Attrition: A Review of Literature. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2868451

Negi, G. (2013). Employee Attrition: Inevitable yet Manageable. *International Monthly Refereed Journal of Research in Management & Technology.* ISSN – 2320-0073, 2.

Parker, Stephen K., & Skitmore, M. (2005). Project management turnover: Causes and effects on project performance. *International Journal of Project Management*, *23*(3), 205–214. https://doi.org/10.1016/j.ijproman.2004.10.004

PLoS Medicine (OPEN ACCESS) Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Powers, D. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Mach. Learn. Technol.*, *2*.

Prabakaran, P., & Vetrivel, D. T. (2017). Conceptual Paper on Workforce Attrition: Causes, Consequences and Prevention. *Star International Journal, 5*(2). ISSN: 2321-676X

Provost, F., Fawcett, T. (2001). Robust Classification for Imprecise Environments. Machine Learning 42, 203–231. https://doi.org/10.1023/A:1007601015854

Punnoose, R., & Ajit, P. (2016). Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. *International Journal of Advanced Research in Artificial Intelligence*, *5*(9). https://doi.org/10.14569/IJARAI.2016.050904

Rhodes, S. R. (1983). Age-Related Differences in Work Attitudes and Behavior: A Review and Conceptual Analysis. *Psychological Bulletin, 93*(2), 328-367.

Ribes, E., Touahri, K., & Perthame, B. (2017). *Employee turnover prediction and retention policies design: A case study*. hal-01556746

Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, *38*(3), 1999–2006. https://doi.org/10.1016/j.eswa.2010.07.134

Selhadin, I. (2019). Causes of Employee Turnover: the Case of Global Insurance Company. [Master dissertation, St. Mary's University]. *St. Mary's University Institutional Repository.* http://hdl.handle.net/123456789/4939

Shapley, L. S. (1951). "Notes on the n-Person Game -- II: The Value of an n-Person Game" (PDF). Santa Monica, Calif.: RAND Corporation. https://www.rand.org/pubs/research_memoranda/RM0670.html

Singh, S. & Singh, S. (2017). Linking Positive Psychological Capital to Work Well-Being and turnover intentions. *International Research Journal Commerce arts science, 8*(1). ISSN 2319-9202.

Staw, B. M. (1980). The Consequences of Turnover. *Journal of Occupational Behaviour, 1*(4), 253–273.

Stephen Milborrow. rpart.plot: Plot rpart Models. An Enhanced Version of plot.rpart., 2016. R Package.

Taylor, M. S., Audia, G., & Gupta, A. K. (1996). The Effect of Lengthening Job Tenure on Managers' Organizational Commitment and Turnover. *Organization Science*, *7*(6), 632–648. https://doi.org/10.1287/orsc.7.6.632

Tuji, A. (2013). An Assessment of the Causes of Employee Turnover in Oromia Public Service Organizations. [MBA dissertation, Addis Ababa University]. https://rb.gy/prcnjf

Varadharaj, A. & Irfan, EC. (2019). A Study Report on Causes and Effects of Employee Turnover in Construction Industry. *International Research Journal of Engineering and Technology (IRJET), 6*(5), 2371-2382. e-ISSN: 2395-0056

Varghese, T. A., Joshi, S., & Sampathkumaran, S. (2019). Attrition risk analyzer system and method (US Patent No. 10,339,483). U.S. Patent and Trademark Office. https://rb.gy/hkicq6

Wang, M.-L., & Chen, W.-Y. (2013). An Exploration of Career Stages from the Causes of Turnover: Nurses as Examples. *International Journal of Management Theory and Practices, 14*(1), 40-54.

Woo, S.E., & Maertz, C. P., Jr. (2012). Assessment of Voluntary Turnover in Organizations: Answering the Questions of Why, Who, and How Much. In N. Schmitt (Ed.), Oxford library of psychology. The Oxford handbook of personnel assessment and selection (p. 570–594). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.013.0025

Workagegn, S. (2017). An Assessment of The Causes of Employee Turnover. A Case of Tikur Abay Shoe S.C. [Master dissertation, St. Mary's University]. *St. Mary's University Institutional Repository.* http://hdl.handle.net/123456789/3502

Zhang, Q., Li, J. & Song, Q. (2018). An Analysis of Relative Structure on Causes of Executive Turnover. *Proceedings of the Third International Conference on Economic and Business Management (FEBM 2018): Vol. 56. Advances in Economics, Business and Management Research* (pp. 374-377). Atlantis Press. https://dx.doi.org/10.2991/febm-18.2018.85

Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2019). Employee Turnover Prediction with Machine Learning: A Reliable Approach. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), *Intelligent Systems and Applications* (Vol. 869, pp. 737–758). Springer International Publishing. https://doi.org/10.1007/978-3-030-01057-7_56

Zou, H. & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society, Series B. 67 (2): 301–320. CiteSeerX 10.1.1.124.4696. doi:10.1111/j.1467-9868.2005.00503.x

Zylka, M. P. (2016). Putting the Consequences of IT Turnover on the Map: A Review and Call for Research. *Proceedings of the 2016 ACM SIGMIS Conference on Computers and People Research*, 87–95. https://doi.org/10.1145/2890602.2890618

Zylka, M. P. & Fischbach, K. (2017). Turning the Spotlight on the Consequences of Individual IT Turnover: A Literature Review and Research Agenda. *ACM SIGMIS Database*, *48*(2), 52-78. http://dx.doi.org/10.1145/3084179.3084185

Sitography

[1] https://smallbusiness.chron.com/employee-turnover-vs-attrition-15846.html

[2] Wigert, B. (2018). Talent Walks: Why Your Best Employees Are Leaving. https://www.gallup.com/workplace/231641/talent-walks-why-best-employees-leaving.aspx

[3] Sullivan, J. (2015). How Commute Issues Can Dramatically Impact Employee Retention. https://www.tlnt.com/how-commute-issues-can-dramatically-impact-employee-retention/

[4] Zhang, A. (2019). Workplace romances: do you need love contracts & non-fraternisation policies? https://www.insidehr.com.au/workplace-romance-love-contracts-non-fraternisation-policies/

[5] Milborrow, S. (2020) Plotting rpart trees with the rpart.plot package. http://www.milbo.org/rpart-plot/prp.pdf

[6] Microsoft Docs: Team Data Science Process. Available online: https://docs.microsoft.com/it-it/azure/machine-learning/team-data-science-process/ (accessed on 31 October 2021).
[7] HR Analytics- exploration and modelling with R. https://www.kaggle.com/ragulram/hr-analytics-exploration-and-modelling-with-r
[8] https://en.wikipedia.org/wiki/Precision_and_recall