UNIVERSITY OF MODENA AND REGGIO-EMILIA
MARCO BIAGI DEPARTMENT OF ECONOMICS

Ph.D. Program in Labor, Development and Innovation
XXXIII Cycle

The Human Side of Digital Revolution: Text Mining Tools to Face Industry 4.0 Phenomenon

SUPERVISORS
Prof. Giovanni Solinas
Prof. Ing. Gualtiero Fantoni

COORDINATOR
Prof. Tindara Addabbo

CANDIDATE
Silvia Fareri

ACADEMIC YEAR 2019/2020

"If it can be lost, then it can be won
If it can be touched, then it can be turned
We said we're going to conquer new frontiers
All you need is time."
Bloc Party - The Pioneers

Ringraziamenti

Quando ho iniziato questo percorso sapevo cosa sarei voluta idealmente diventare, ma non sapevo come riuscirci; mi sentivo priva, più che dei mezzi, di una strada tracciata da percorrere. Dopo tre anni la situazione è radicalmente cambiata e per questo devo ringraziare diverse persone.

Ringrazio il Professor Fantoni per aver scommesso su di me fin dal primo istante, o meglio, fin dal primo conflitto (e tutti gli innumerevoli a seguire); ce ne auguro almeno altrettanti. Ringrazio il Professor Solinas che mi ha offerto immediatamente grande disponibilità, concreto supporto e mi ha mostrato un modo radicalmente diverso di vedere (e vivere) le difficoltà; i vostri due approcci sono stati ortogonali, e non avrei potuto immaginare combo migliore per la mia crescita personale.

Ringrazio Diego e la Fondazione Giacomo Brodolini, per aver investito su di me e per continuare, ancora adesso, ad alzare l'asticella, a pormi davanti obiettivi che pensavo essere irrealistici, ma che così non sono mai stati.

Ringrazio il mio team di ricerca (in particolare Elena, Nicola, Pietro e Vito), i ragazz* di Erre Quadro e del GATE, per avermi dato la possibilità di trovare non solo dei colleghi, ma anche dei veri amic*; non vi nominerò tutti, ma voi sapete.

Ringrazio Filippo e il nostro esclusivo modo di ascoltarci, parlarci e comprenderci intimamente (spesso silenziosamente).

Ringrazio I miei colleghi di Modena (in particolare Stefano, Luca, Andrea, Federica, Giorgio), con cui ho vissuto questo percorso e che hanno immediatamente azzerato la sensazione di estraneità che ho percepito I primi momenti; mi voltavo e avevo la certezza di non essere sola, mai.

Ringrazio I miei migliori amici, Nicola e Matteo, I miei più grandi motivatori e parallelamente chi mi ha dato più modo di mettermi in discussione.

Ringrazio I miei amici di Pisa e Spezia e chi ha vissuto con me i mesi paradossali della Quarantena; legami indissolubili, per me.

Ringrazio le mie amiche di sempre (Giulia, Clara, Ilaria, Andrea e Linda), i miei punti fermi, le mie certezze; se sono così, lo devo soprattutto a voi.

Ringrazio Antonio, per la naturalezza, purezza e irrazionalità di ciò che ci lega e che mi ha permesso di vivere l'ultimo anno di questo viaggio complesso con una serenità raramente provata in vita mia.

Ringrazio infine I miei genitori, di cui rappresento la perfetta fusione di pregi e difetti; spero che siate fieri di ciò che la me adulta è diventata e si appresta a diventare.

Auguro, infine, a chi ha coscienza di ciò che vorrebbe essere ma non conosce la strada, che sia il più tortuosa, faticosa e soprattutto il più ripida possibile... la vista, da qui, è davvero magnifica.

Acknowledgments

When I first set out on this journey, I ideally knew what I wanted to become, but lacked the means to achieve my goal. Today, this has changed radically, and I would like to express my sincerest gratitude to several key actors.

I would like to thank my esteemed Professor Fantoni for his invaluable supervision, and for having believed in me from the very first moment (... or should I say conflict) and all the ones that followed; hopefully, they will continue to follow in the coming future. I would like to offer my special thanks to Professor Solinas for his continuous support, availability, and for showing me a totally different way of facing and dealing with academic (and personal) problems. Two complementary approaches that have been priceless in my personal and professional growth, and for this reason I sincerely thank you.

I would also like to express my gratitude to Diego and the Giacomo Brodolini Foundation, for having invested in me and for continuing to do so by setting new challenges and objectives, that seemed to be often unrealistic, but actually they have never been.

A very special thanks to my research team (in particular, Elena, Nicola, Pietro and Vito) and the "guys" from Erre Quadro and Gate for giving me the chance to meet and work not only with colleagues but real friends. I will not mention everyone here, but you all know who you are and what you have meant for me.

I would also like to thank Filippo and our exclusive way of listening and understanding each other, often in meaningful silence.

My deepest gratitude to my colleagues from Modena (in particular Stefano, Luca, Andrea, Federica and Giorgio) with whom I shared this experience and who welcomed me from the very first day, making me feel immediately at home.

I thank my best friends, Nicola and Matteo, my greatest motivators and supporters and at the same time whom ask questions of me the most.

I would also like to thank my friends in Pisa and Spezia and those with whom I shared the months of the Covid-19 lockdown; an indissoluble bind for me.

My sincerest thanks to my best friends (Giulia, Clara, Ilaria, Andrea and Linda) who have always been there for me with their loyal support and love and helped me become the person I am today.

I thank Antonio, for the naturalness, innocence and irrationality of our relationship, which have helped me get through this last challenging year.

I would like to thank my parents whose strengths and weaknesses I embody. I hope you are proud of the woman I have become and will continue to become along this journey of life.

Finally, to those who are aware of what they would like to be but do not know the way, I hope it it will be as tortuous, as tiring, as steep as possible... the view for up here is breathtaking.

Abstract

The advent of Big Data and their progressive increase in volume and complexity has allowed the development of completely new scenarios for companies. Particularly, the techniques of management, analysis and conversion of raw data into "sexy" information are constantly evolving. At the same time, the increasingly widespread paradigm 4.0 discloses radical changes not only in technology but also in the structure and dynamics of work, changing skills and abilities required by manufacturing and services. In addition, if routinely tasks seem to be susceptible to digitalization, on the other hand ownership of transversal skills (also known as soft skills) is ever more recognized as a bottleneck for automation and they are ever more required by the labour market.
In view of this, the purpose of the present work is to study the effects of Industry 4.0 on the workforce, developing ad hoc Text Mining tools to properly manage its impact, and providing a tangible support in this transitional and critical phase. Moreover, this thesis focuses on satisfying the needs of three main stakeholders: HR management in recruitment, reskilling and upskilling; Institutions to build customized learning paths and to update International recognized databases; Policy makers to foresight the heterogeneous effects of digitalization on job profiles, analysing the change of skill demand on supply chains in Emilia-Romagna.

L'avvento dei Big Data e il loro progressivo aumento di volume e complessità ha permesso lo sviluppo di scenari completamente nuovi per le aziende. In particolare, le tecniche di gestione, analisi e conversione dei dati grezzi in informazioni "sexy" sono in continua evoluzione. Allo stesso tempo, il paradigma 4.0 sempre più diffuso svela cambiamenti radicali non solo nella tecnologia ma anche nella struttura e nelle dinamiche del lavoro, modificando competenze e capacità richieste dalla manifattura e dai servizi. Inoltre, se le attività di routine sembrano essere suscettibili alla digitalizzazione, d'altra parte il possesso di competenze trasversali (note anche come soft skills) identifica sempre più un collo di bottiglia per l'automazione, e le stesse risultano essere sempre più richieste dal mercato del lavoro. In relazione a ciò, lo scopo del presente lavoro è studiare gli effetti dell'Industria 4.0 sulla forza lavoro, sviluppando strumenti di Text Mining ad hoc per gestirne adeguatamente l'impatto e fornendo un supporto tangibile in questa fase critica di transizione. Inoltre, questa tesi si concentra sul soddisfare le esigenze di tre principali stakeholder: la gestione delle risorse umane nel recruiting, riqualificazione e miglioramento delle competenze; le Istituzioni per costruire percorsi di apprendimento personalizzati e aggiornare database riconosciuti come standard a livello internazionale; i policy maker per prevedere gli effetti eterogenei della digitalizzazione sui profili professionali, analizzando il mutamento della domanda di competenze sulle filiere in Emilia-Romagna.

Contents

1 Introduction

New technologies, new paradigms, new skills. An industrial revolution represents a radical change in socio-economic and political systems driven by the introduction of new methodologies and technologies that cause a significant increase in efficiency and productivity (Tarry, 2019). For what concerns the fourth one, also known as Industry 4.0, existing literature shows many heterogeneous attempts in defining it; to our knowledge, its impact is certain but still not defined yet (Last, 2017). The more specific definitions focus on the interconnections between technologies: for instance, Lasi et al. (2014) define I4.0 as the increase of digitalization and automation of the manufacturing scenario, as well as the creation of digital value chains to make products, environment and business partners able to communicate with each other. A similar definition is also given by Oesterreich and Teuteberg (Oesterreich and Teuteberg, 2016). Conversely, there are also more general definitions, as for example those provided by Hermann et al. (Hermann et al., 2016), who define I4.0 as a collective term for technologies and concepts of value chain organization, and by Szalavetz (Szalavetz, 2018), who addresses I4.0 as a sort of umbrella term for identifying a variety of digital enterprise technologies.

Despite these given definitions, many researchers (Arnold et al., 2016; Porter et al., 2014) outline that the Industry 4.0 phenomenon does not only interest manufacturing practices, but it has a broader scope, also affecting socio-economic, geopolitical and demographic developments. That being said, the new paradigm cannot be merely considered a technological tsunami, but it is made by heterogeneous elements and may have heterogeneous consequences also from a managerial and governmental point of view (Arnold et al., 2016; Porter and Heppelmann, 2014). Regarding the effects of I4.0 on labor market, it is possible to find different schools of thoughts. On one hand, the automated systems, the backbone of paradigm 4.0, are considered a sort of threat: some professional jobs could be vulnerable, and they are likely to be substituted, as Artificial intelligence and Big data give machines ever more human-like abilities (Rotman, 2013). Furthermore, although technologies could enhance the productivity of many workers, they seem to lead to an employment of the workforce in 'bad jobs' (Lloyd and Payne, 2009) and low skilled workers are reported being stressed and constantly under pressure (Lafer, 2004). For what concerns the likelihood of worker's substitution by machines, the research of (Frey et al., 2017) is emblematic: they estimate around 47% of jobs is in the high-risk category, especially the ones characterized by routine tasks.

Conversely to Frey's point of view, Caruso (Caruso, 2017) states that technological innovation is not substituting less-skilled workers, but up to now, it is producing results that are always being achieved in history of capitalism, such as increasing efficiency. Dengler et al. (Dengler et al., 2018) also outline that previous works have overestimated Industry 4.0 impacts on the labor market, in particular the suscepti-

bility to automation of occupations. When they assumed that entire job profiles are replaceable, they obtained results similar to the ones obtained in previous studies; but when they considered that only certain tasks would be substituted, they estimated that 15% of German employees are actually at risk. To her same school of thought belongs Rosenberg (Rosenberg, 2009), who views the concept of automation not as a threat but as an opportunity: human workers would be probably free to express their talents, focusing their energies where they could add value. According to this, the new technologies could have a positive impact on employment, more specifically 3D printing, Internet of Things, Augmented reality and Big data analytics demand a large quantity of new skills to be properly managed (Freddi, 2017), thus increasing human labour. Following this idea, MacCrory (MacCrory, 2014) hypothesizes three main consequences of technological innovation, which can be summarized in: a significant reduction in skills that compete with automation; a significant increase in skills which complement machines; finally, an increase in skills where machines are not advanced enough. Regardless the numerous definitions and studies carried out on the paradigm, it is certain that the latter is growing at an exponential rate and its impact on the industries, and not only, is unpredictable.
Focusing on skills, more than 50 years ago, the pioneering work carried by the US Bureau of Labor Statistics (BLS) made governments in most developed economies aware of the strong need for assessing national labour market prospects and competencies projections (Wilson, 2013). Today, the proactive adaptation which leaders must begin to consider, also embraces the need to manage skill requalification and disruption. From a managerial point of view, the presence of the wrong skills mixes leads enterprises to have bad performances (Grugulis et al., 2009; Lorentz et al., 2013). However, as organizations are complex political entities, determining the right mix of skills for a firm is not an easy task (Abbott, 1993).
Within this framework, many works are specifically aimed to define which are the characteristics that could make a job profile resilient to change. (Chryssolouris et al., 2013; Gorecky et al., 2014; Weber, 2016). Linked to this, other studies (Bauer et al., 2011; Bridgstock, 2011; Dobrunz et al., 2006; Haukka, 2011; Cooper and Tang, 2010) introduced the criticality of competence gap both in companies and universities and their consequent central role.
Instead, (Kamprath et al., 2015) focus on companies, discussing how the changing business environment is affecting working conditions and competences of individuals, also identifying the future request of skills. At the same time, Academia and industry have shown a growing interest on soft skills, particularly in the last few years. The focus of scholars and practitioners on this topic has grown for many reasons, but the main one is (counterintuitively) digitalisation. It is known that the impact of digitalisation is widespread (Van Laar et al, 2017) and heterogeneous (Galati et al., 2017), but is also evident that many studies outline the importance of acquiring soft skills in order to face the digital wave (Chryssolouris et al., 2013;

Gorecky et al., 2014; Weber, 2016), stressing that a competence gap in this context will have a negative impact on the workforce.
In addition, several years ago (Acemoglu et al., 2010; Levy et al., 2004) stated that machines can replicate what is easily encoded, whereas, concepts which are not definable with rules cannot be understood. This is still true and relevant, since job roles and responsibilities are changing and their routinely tasks automated (Frey, 2017) but soft skills seem to represent a digitalisation bottleneck. Moreover, the lack of a universal definition of soft skills domain in contemporary literature is a proof of fuzziness of the concept.

In light of this, it is increasingly necessary to design tools and methods to spot talents and skill gaps: the insights are useful for several objectives and they help organizations to maximize their profits and to improve their HR management strategies (World Economic Forum, 2016). For an effective application of the technologies, it is strictly necessary to carefully observe company preconditions. The latter have to be properly identified in order to integrate the bases of Industry 4.0 (Crnjac et al., 2017).
As we can infer from previous works, there have been many attempts to implement methodologies for managing the technological changes of I4.0. New methods for HR management in the context of I4.0 are clearly rarer, maybe because the impact on the labour market is definitely more complex to be measured and because the phenomenon is relatively new. However, new methods and techniques, such as Data Mining and Analytics, represent a fundamental support to reach the previous goal. In Europe, skill forecast is one of the main issues carried on by the European Centre for the Development of Vocational Training (Cedefop). By using harmonized international data and common approaches for European countries, it provides every year not only quantitative projections of the future trends in employment, but also cross-country comparisons. Other works start from data to predict the future of labour market. Among these, the author can mention the report “Skills Needs Analysis for Industry 4.0 Based on Roadmaps for Smart Systems” (HartmannBovenschulte, 2013) in which, starting from some roadmaps linked to Smart Systems, skills demand is derived and quantitatively explored by constructing scenarios and technology/-sector matrices. In the study “Technology usage, expected job sustainability, and perceived job insecurity” (Nam, 2018), data analysis supports the prediction of job insecurity in United States, starting from data collected by telephone survey.

In this context, a key source of information is represented by the taxonomies of skills, which embody the standard language of professions and skill requirements. Several of them can be found online, but the primary sources of occupational information are ESCO[1] (European) and O*NET[2] (American). ESCO (European Skill/Competence Qualification and Occupation) is a multilingual classification system for Europe; it classifies jobs, capabilities, competences, and qualifications in Europe that are relevant for the labor market. Through a triangular relationship among skills, profiles, and qualifications, the aim of ESCO is bridging the gap between academia and industry in all of Europe. The occupation classification corresponds to ISCO-O8, which is the International Standard Classification of Occupations (International Labor Organization, 2008). O*NET is the American correspondent of ESCO. O*NET is an available online database developed for the U.S Department of Labor which is made of 974 occupations from Standard Occupational Classification (SOC) and their corresponding skills, knowledge, and abilities. The use of SOC makes it possible to analyse professions from multiple perspectives, comparing data from different federal sources, aggregating data on employment from a wide range of job titles, and tracking data over time to identify changes in the labor market. Each job profile has quantitative information about level and importance for every owned skill described above. ESCO also has a greater level of detail than O*NET, having 6 times the number of skills and 3 times the number of job profiles (of O*NET). Furthermore, ESCO contains a large number of heterogeneous skills, which are frequently assigned only to a single job profile. Finally, there is not clear distinction between hard and soft skills in O*NET, while around 110 skills are labelled as transversal[3] in ESCO (v1.0.3); however, also ESCO has some weakness, because its soft skills are too abstract to be assigned to a specific job profile[4] and their number is still limited. The two databases presented above embody the state-of-the-art of skills and job profiles and they must be properly studied. However, they are a static representation, a photograph, and they could not capture a highly dynamic phenomenon such as the evolution of the needs of the labour market.

In relation to the above, the adopted methodology is founded on the use of data science and on the implementation of state-of-art algorithms, which allowed the continuous analysis of a huge amount of textual data from heterogeneous sources, including the skills frameworks themselves. The methodological framework is basically founded on data-driven approaches using Natural Language Processing (NLP), and in particular Named Entity Recognition (NER) Tools (Harrag, 2014).
Information extraction from unstructured documents is a task that has been subject of active research for several years in the community of Artificial Intelligence (Harrag, 2014) and Computer Engineering (Choudhary et al., 2009). Current approaches focus on processing text in order to make a meaningful representation of

[1]https://ec.europa.eu/esco/portal/home
[2]https://www.onetcenter.org
[3]File transversalSkillCollection.csv
[4]https://ec.europa.eu/esco/portal/document/it/87a9f66a-1830-4c93-94f0-5daa5e00507e

concepts (and their relationships) contained in documents (Piskorski, 2013; Gildea, 2014; Liu et al., 2020). In particular, NER consists of detecting lexical units in a word sequence that refers to a predefined entity, thus determining what kind of entity the unit is referring to (e.g. persons, locations, organizations.). The methods used for NER are various:

1. terminological-driven NER: aims to map mentions of entities within texts to terminological resources (e.g. wikipedia) (Nadeau et al, 2017);

2. rule-based NER: uses lexicons, regular expressions and lexical information to express knowledge based systems able to extract a certain type of entity (Sari et al, 2010);

3. corpus-based NER: uses manually tagged text corpora (training set) to train machine learning (ML) algorithms (Quinlan, 1986; Suykens, 1999; Lafferty et al, 2001; McCallum et al, 2010)

NER has been successful in different languages and different domains. It can provide crucial, although shallow, semantic information for tasks such as question answering (Abujabal et al, 2018; Blanco-Fernández et al., 2020), topic disambiguation (Fernández, N. et al. 2012) or detection (Krasnashchok et al, 2018; Lo et al., 2017, Al-Nabki et al., 2019) and revealment of elements relationships (Sarica et al., 2020; Amal et al., 2019). Furthermore, NER has proved to be effective in broader applications, such as user profiling (Nicoletti et al., 2013) and ontology development on unconventional domains (Oliva et al., 2019; Rodrigues et al., 2019) and for these reasons it is the main technique applied.

To conclude, the goal of this project is represented by the development of Text Mining tools (scalable and adaptable) which should be customizable to meet multiple and heterogeneous needs such as: providing precious guidelines for HR management, measuring the current impact of digitalization on internal job profiles; policy makers and Istitutions to face Industry 4.0 requirements, according to their relevance and impact on employability; updating international recognized databases of skills collecting real-time insights from reliable sources. Since the phenomenon is quite complex and its boundaries are definitely fuzzy, it was necessary to define the impact of digitalization on jobs and skills considering different perspectives. For doing so, the methodological framework is based on heterogeneous analytic approaches and, coherently, the thesis covers multiple research areas, such as Human Resource Management, Industrial Economics and Innovation, which is one of the main distinctive elements of the present work.

The thesis consists of three chapters.

On Chapter 1, the author tried to develop a measure for quantifying the readiness of employees belonging to a big firm with respect to the Industry 4.0 paradigm. To reach the goal, a data-driven approach based on text mining techniques was applied to a case study. In particular the present methodology made use of a previously developed enriched dictionary of technologies and methods 4.0 (Chiarello et al., 2018). The source was used to analyze job profiles' descriptions belonging to Whirlpool, a multinational company with a structured database of jobs and skills. The process allowed the identification of technologies, techniques and related skills contained in job descriptions. Starting from these, the Industry 4.0 impact on each job profile was measured. Finally, the metadata of the job profiles were analyzed to evaluate to which extent the skills of profiles 4.0-ready and non-4.0-ready differ. In the end, the work provided a framework for estimating the Industry 4.0 readiness of enterprises' human capital which demonstrated to be fast, adaptable and reusable.

On Chapter 2, the author had the aim of developing an automated tool capable of extracting soft skills from unstructured texts. Starting from an initial seed list of soft skills, the author automatically collected a set of possible textual expressions referring to soft skills, thus creating a Soft Skills list. This has been done by applying Named Entity Recognition (NER) on a corpus of scientific papers developing a novel approach and a software application able to perform the automatic extraction of soft skills from text: the SkillNER. The author measured the performance of the tools considering different training models and validated the approach comparing the obtained list of soft skills with the skills labelled as transversal in ESCO (European Skills/Competence Qualification and Occupation). Finally, the author showed a first example of how the SkillNer can be used, identifying the relationships among ESCO job profiles based on soft skills shared, and the relationships among soft skills based on job profiles in common. The final map of soft skills-job profiles may help accademia in achieving and sharing a clearer definition of what soft skills are and fuel future quantitative research on the topic.

Finally, on Chapter 3, starting from the analysis of turnover data, the author analyzes the trends in the demand for skills in Emilia-Romagna in the decade 2008-2017. The research was conducted through the analysis of the SILER database (Emilia-Romagna Labor Information System), in which the survey unit is represented by the movements of hired and fired persons resulting from the mandatory communications of employers. The investigation is divided into two phases. Initially, particular attention was paid to the trend of hires, and to the net balance of hires, highlighting the professions most requested by the different supply chains and, on the other hand, the ones that are less and less relevant. In the second phase, the author performed a crosswalk between CP2011 code (classification of ISTAT professions) and the European classification of skills and professions, with the aim of providing a picture of the evolution of the demand of skills. At this stage, through the aid of a previously developed dictionary of methods and technologies (Chiarello et al, 2018), it was also possible to identify the main skills 4.0 related to the most requested professions.The analysis and instrumentation proposed in the last essay could represent an useful tool for monitoring company choices and building coherent and adequate industrial policy (and training) tools.

In a rapidly evolving scenario following Industry 4.0, the shock caused by the Covid-19 emergency will have a further impact with repercussions not only on employment but also on the organization of production processes and on the demand of new skills. In relation to the above, and the main conclusions of the following three works, the author eventually proposes a positive view of digitalization phenomenon, that should be considered even more an opportunity and will have a fundamental role in this further and critical historical phase.

2 Estimating Industry 4.0 Impact on Job Profiles and Skills using Text mining[5]

2.1 Introduction

Today there is an urgent need to understand which is the impact that technology is having on the workforce. This urgency is given by the paradigm of Industry 4.0: a sociotechnical revolution, having repercussions on Human Resources as well as on Technological Resources. Industry 4.0 is a systemic transformation in manufacturing and economy which is also influencing society, governance and human identity (Sung, 2018). The new paradigm cannot be merely considered a technological tsunami, but it is made by heterogeneous elements and may have heterogeneous consequences also from managerial and governmental point of view (Arnold et al., 2016; Porter and Heppelmann, 2014). In particular, this impact is affecting multiple groups of stakeholders: companies that have invested in digital innovation in the last 5 years (since the advent of Industry 4.0) are now in the need for an alignment of their internal competencies to maximise the return on investments; labour force feels threatened by robots and Artificial Intelligence which are succeeding in many new tasks; governments are trying to look to the future of sectors that characterize modern economy; universities are reshaping their offer almost every year. The understanding of changes taking place is increasingly crucial for the whole society since a more detailed knowledge of skills requirements helps in designing training programmes giving the opportunity to upskill and reskill (Cedefop, 2019). Despite the great effort of these stakeholders, there exists a lack of tools able to detect the impact that technology is having on specific jobs. This is mainly due to the fact that technological change is becoming increasingly challenging to measure, given its rapid (and often unpredictable) consequences on the worldwide economy. The speed of change is growing as well as global competition in many sectors, with outgoing players and new players. The dynamicity of the landscape creates the needs for updated maps of players on the market and their assets. Big data promises to give answers to these questions, but every time there is the intent of collecting new knowledge about technological innovation, dealing with the velocity, veracity, variability and variety of data is a major issue (McAfee, 2012). On the other side, skill anticipation has been carried out for decades now and the studies performed by different global, european and national actors provide an in-depth picture of the main phenomena happening around the world (Bakule et al., 2016; FGB, 2017; OECD ILO, 2018; CEDEFOP, 2019;CEDEFOP, 2016). Unfortunately, although they provide useful

[5] *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.*

insights at macro level, they are not sector specific and they are far from being company specific. Therefore companies struggle in reshaping their skill inventory, job profiles and organisational structure, and they are forced to use quick-fixes, non-reproducible approaches, usually by using a trial and error approach. Since human capital represents the most valuable asset for an organization (Fulmer et al., 2013), firms could derive extremely competitive advantages from its valorization (Barney, 1991; Becker et al., 1996; Lado et al., 1994). In view of the above, identifying the impact of new technologies on human resources is becoming ever more necessary and strategic. Thus, we formulated the research question of the present paper as follows: "how to develop a semi-automatic procedure to estimate the Impact of Industry 4.0 on Job Profiles and Skills?". In the present work we thus developed a system based on a state of the art technologies mapping tools (Chiarello et al., 2018) in synergy with Natural Language Processing techniques in order to automatically estimate the impact of Industry 4.0 on job profiles. With the purpose of validating the method and showing which output can be obtained, the tool has been tested on the internal documentation of an Italian plant of the multinational company Whirlpool. With respect to the job descriptions of Whirlpool, we have three different outputs: the job profiles that are impacted by Industry 4.0 and their characteristics; a ranking of the most frequent (thus, the most acquired) technologies 4.0; a graphic representation of the transversality of skills 4.0 in relation to the heterogeneity of their correspondent job profiles.

The paper is structured as follows: firstly, a literature review is carried out to better understand both the changes I4.0 will bring and the potential applications of text mining techniques in the field of Human Resources. Secondly, the methodology adopted to reach the analysis objective is deeply described. In the end, some examples of extractions and visualization of the results are shown, followed by the discussion of the results, implications and conclusions.

2.2 Industry 4.0 Phenomenon and its labor market impact

2.2.1 The rise of Industry 4.0

An industrial revolution represents a radical change in socio-economic and political systems driven by the introduction of new methods and technologies that cause a significant increase in efficiency and productivity. (Tarry, 2019) More specifically, the literature defines three main industrial revolutions that preceded the digital one, even if there are still disputes about their dating, number and heterogeneity (von Tunzelmann, N., 2003). Conventionally, the First one took place in Britain at the end of the 18th Century with the introduction of water and steam into production processes and the use of a new material (the iron) and new machineries (Thoben, 2017; von Tunzelmann, N., 2003). The Second Industrial revolution started in

Germany and America between 1870 and 1914, and introduced mass production in assembly lines, the use of electricity and oil and the creation of steel and plastic. Finally, the third one took place and emerged in the end of the 20th century in a multitude of industrialised countries. It consisted of the automation of processes through ICT and biotechnology, with the birth of silicon and smart materials (Thoben, 2017; von Tunzelmann, N., 2003). The rise of the Fourth Industrial Revolution started in 2011, when the term "Industry 4.0" was firstly used during the Hannover Fair. Since that moment, the research interest in I4.0 has started growing as shown in Figure 1. To count the number of scientific papers on the topic of I4.0 we searched on Scopus Database for the following query: "second machine age" OR "fourth industrial revolution" OR " digital age" OR "digital economy" OR "industry 4.0" . The trend is clearly exponential.

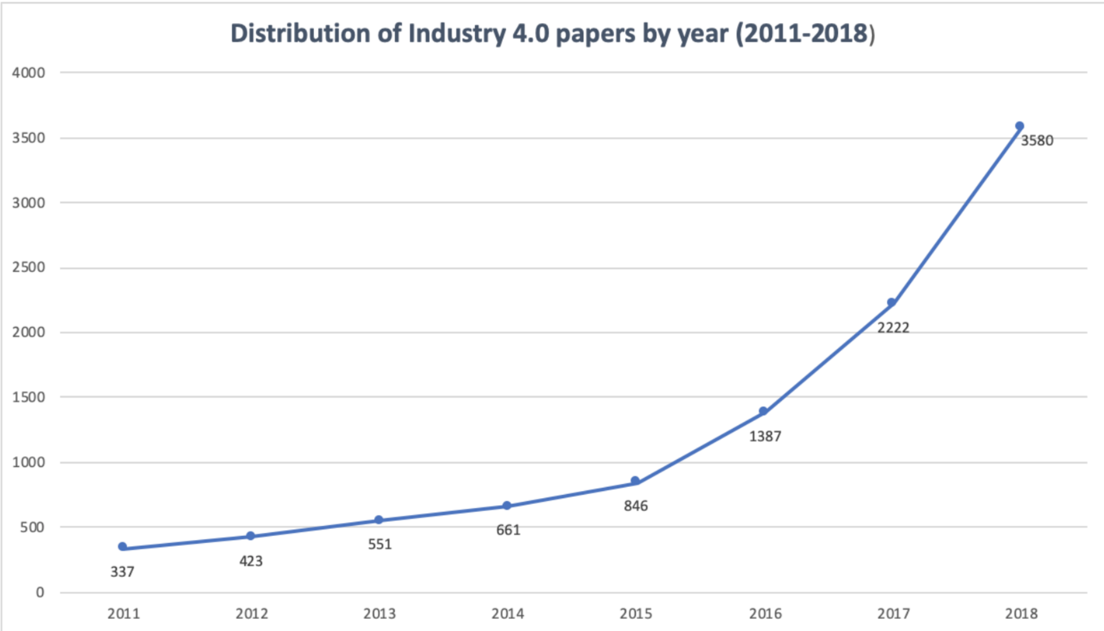


Figure 1: Distribution of Industry 4.0 papers by year (2011-2018)

To have an even deeper understanding of the growing of interest on I4.0, Figure 2 shows the distribution of I4.0 papers among some relevant countries starting from 2011. Curiously, Germany, that is supposed to be the first mover in the I4.0 field, published a small number of papers in 2011, if we compared it to the United States; anyway, German papers have the most rapid growing trend and Germany reached the higher number of publications in 2018. For what concerns Italy, the research interest of this country in I4.0 started growing only in 2016 and Italy overcame both China and the United Kingdom in number of published papers in 2018.

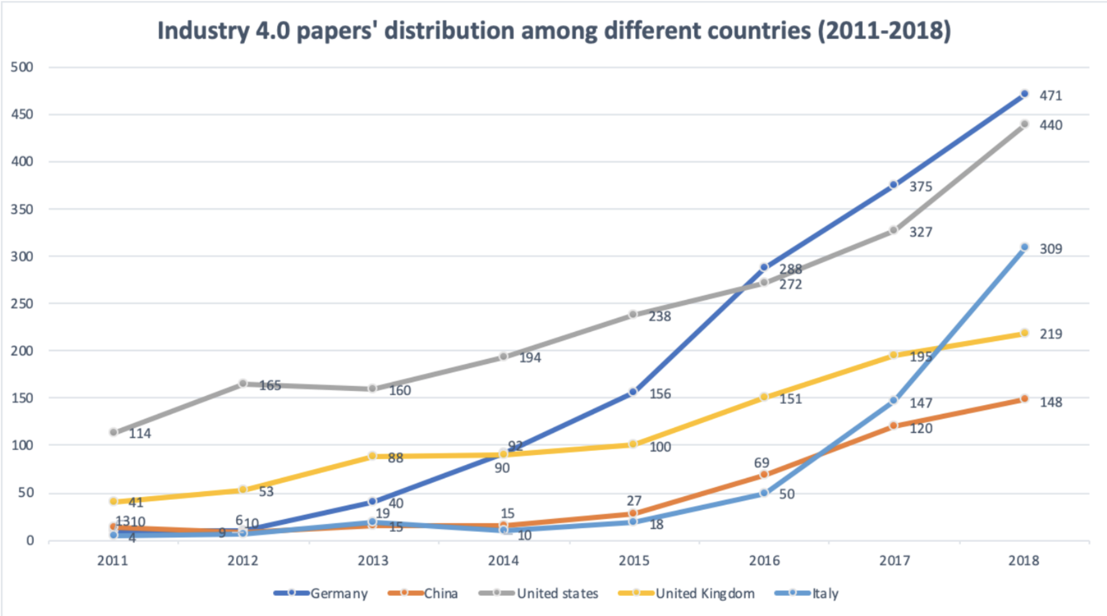


Figure 2: Distribution of Industry 4.0 papers among different countries

2.2.2 Industry 4.0 definitions

I4.0 is usually renamed as Smart Factory, but also embodied in the concept of Cyber-Physical Systems (CPS) (Glas Kleemann, 2016); as we can see, existing literature shows many heterogeneous attempts to define I4.0 (Last, 2017). Moreover, the definitions could be more or less specific. The more specific ones focus on the interconnections between technologies: for instance, Lasi et al. (2014) define I4.0 as the increase of digitization and automation of the manufacturing scenario and the creation of digital value chains to make products, environment and business partners able to communicate with each other. Also Oesterreich and Teuteberg offer a specific definition of I4.0, describing it as an increase of manufacturing environment's digitization, automation and communication enabled by the creation of a digital value chain (Oesterreich and Teuteberg, 2016). Conversely, more general definitions can be found in literature, as for example those provided by Hermann et al. (Hermann et al., 2016), who define I4.0 as a collective term for technologies and concepts of value chain organization, and by Szalavetz (Szalavetz, 2018), who defines I4.0 as a sort of umbrella term for identifying a variety of digital enterprise technologies. Moreover, many researchers (Arnold et al., 2016; Porter et al., 2014) outline that the new revolutionary elements cannot be restricted to enhance production parameters driven by new technologies acquisitions. According to this point of view, concurrently to the technological drivers of change, the broader socio-economic, geopolitical and demographic developments will have almost the same key role and they have to be carefully considered.

2.2.3 Industry 4.0 labor market impact

For what concerns the effects of I4.0 on labor market, it is possible to find different schools of thoughts. On one hand, the automated systems, the backbone of paradigm 4.0, could be considered as a threat: although they could enhance the productivity of many workers, some professional jobs could be vulnerable and they are likely to be substituted, as Artificial intelligence and Big data give machines more human-like abilities (Rotman, 2013). Furthermore, skills-related problems lead to an employment of the workforce in 'bad jobs' (Lloyd and Payne, 2009) and low skilled workers are reported being stressed and constantly under pressure (Lafer, 2004). For what concerns the likelihood of worker's substitution by machines, the research of Frey et al. is emblematic: they tried to define which would be the jobs that will resist and which not, according to the likelihood of computerization of individual's skills (Frey et al., 2017). The results are quite pessimistic: they distinguish between High, Medium and Low risk of computerization, and, according to their estimates, around 47% of jobs is in the high-risk category, especially the ones characterized by routine tasks. Conversely to Frey's point of view, Caruso (Caruso, 2017) states that technological innovation is not substituting less-skilled workers, but up to now it is producing results that are always being achieved in history of capitalism, such as increasing efficiency and reducing workforce and wages. Dengler et al. (Dengler et al., 2018) also outline that previous works have overestimated Industry 4.0 impacts on the labor market, in particular the automation probabilities of professional profiles. When they assumed that entire occupations are replaceable, they obtained results similar to the ones belonging to previous studies; but when they considered that only certain tasks would be substituted, they estimated that 15% of German employees are actually at risk. The same school of thoughts belongs to Rosenberg (Rosenberg, 2009), who sees the automation increase not as a threat but as an opportunity: human workers would be probably free to express their talents, focusing their energies where they could add value. According to this, the new technologies could have a positive impact on employment. The reason why that would probably happen is that 3D printing, Internet of Things, Augmented reality and Big data analytics demand a large quantity of new skills to be properly managed (Freddi, 2017). Following this idea, MacCrory (MacCrory, 2014) hypothesizes three main consequences of technological innovation, which can be summarized in: a significant reduction in skills that compete with automation; a significant increase in skills which complement machines; finally, an increase in skills where machines are not enough advanced.

2.3 Competences and Job Profiles in Industry 4.0

Many works are aimed to define Industry 4.0 effects on employees, particularly focusing on job profiles that require new skills in order to be resilient to change. (Chryssolouris et al., 2013; Gorecky et al., 2014; Weber, 2016). In connection to this, other works (Bauer et al., 2011; Bridgstock, 2011; Dobrunz et al., 2006; Haukka, 2011; Cooper and Tang, 2010) introduced the criticality of competence gap both in companies and universities and their consequent central role. (Cacciolatti et al., 2017) argue that universities need to find the proper position in the knowledge economy; they suggest that universities, policy makers and firms have to collaborate and create programs to obtain practical experience and enable students to learn both hard and soft skills. Instead, (Kamprath et al., 2015) focus on companies, discussing how the changing business environment is affecting working conditions and competences of individuals, also identifying which would be the skill needed in the future. Moreover, different works show the need to clarify the concept of soft skills and the way to teach and improve them. It happens because soft skills seem to be the new essential for the workforce of the future, and not least the key to distinguish men from machines, as literature identifies the most soft-skilled profession as resilient (Frey et al., 2017). Regardless of the point of view, and the consequent heterogeneous needs, Industry 4.0 and its impact on the labor market should be properly studied. The approaches and sources used to study the latter topic are heterogeneous and the authors try to summarize them in Figure 3.

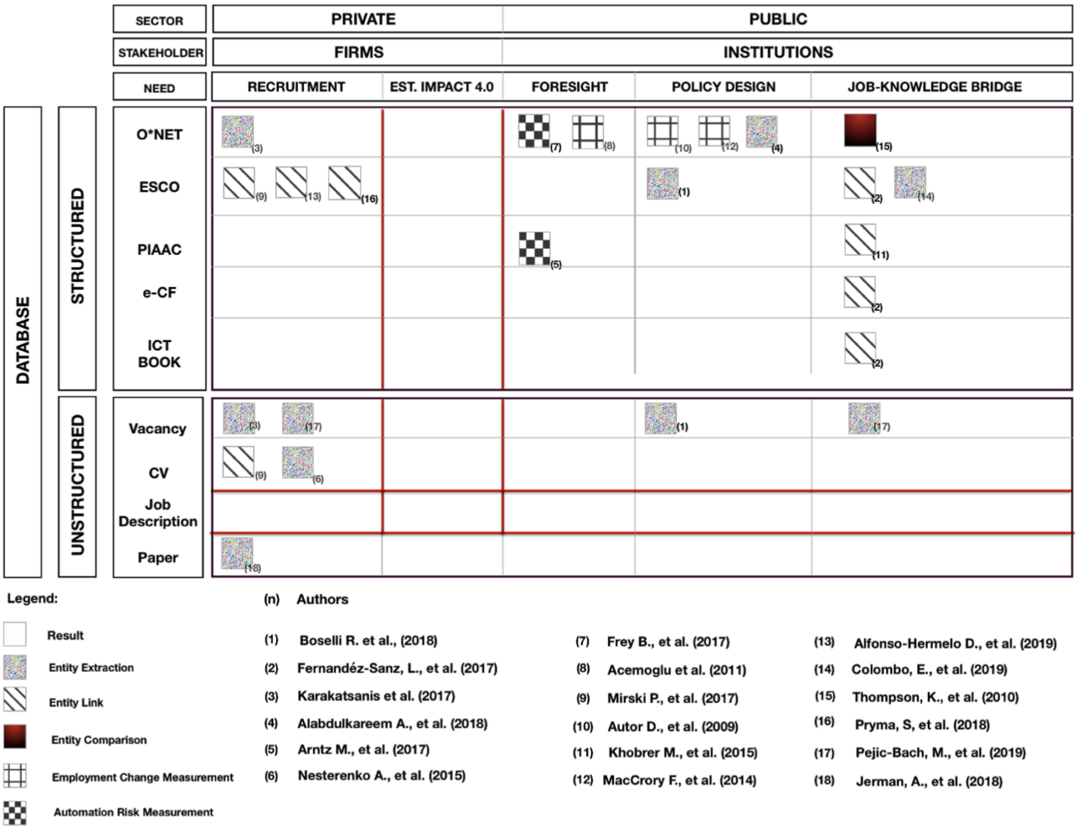


Figure 3: Skill Literature Map, representation by authors. The map should be read as follows: the author "n" responds to the need of the stakeholder "k", who belongs to sector "m", offering "p" as the main result of his analysis and as a possible support/response to his need, analyzing the database "o".

Figure 3 shows the main results of the most cited literature, analyzing heterogeneous databases to satisfy the new demands of the labor market. Starting from the sources, the main databases found in literature can be divided in two macro-classes: structured and unstructured. With the term structured, authors identify the internationally recognized databases of competences and job profiles characterized by a well-defined and unchanging framework, while the term unstructured identifies non-conventional sources. The structured databases considered are:

- ESCO (European Skill/Competence Qualification and Occupation), which classifies jobs, capabilities, competences and qualifications in Europe. Through a triangular relationship among skills, profiles and qualification, the aim of ESCO is bridging the gap between academia and industry in all Europe (European Commission, 2013);

- O*NET, which is an available online database developed for the U.S Department of Labor which is made by 672 occupations and their corresponding skills, knowledge, and abilities; the O*NET database is updated regularly[6];

- PIAAC, (The Program for the International Assessment of Adult Competencies) made by OCSE, which contains measures about adults' proficiency in key information-processing skills - literacy, numeracy and problem solving in technology-rich environments [7];

- The ICT BOK (Body of Knowledge) made by CEPIS is the accepted ontology for a specific domain (ICT skills) and it is defined essentially as a BoK repository for the professional knowledge [8];

- e-CF is the European e-Competence Framework, which provides a reference of 40 competences in Information and Communication Technology (ICT) workplace, expressed through a common language for competences, skills, knowledge and proficiency levels [9].

On the other hand, the main potential unstructured and valuable databases, partially analyzed in literature, are: Curriculum Vitae (Mirski et al., 2017; Nesterenko et al., 2015), Job Vacancy (Karakatsanis et al., 2017), Job Description and Scientific Papers. The first three sources represent current demand and supply of competences and they are an accurate picture of what labor market is both offering and asking for. Scientific papers show where the research is focusing, spotting where new competences can emerge due to technological and scientific advancements. Because of their reliability and usability, O*NET and ESCO are the most analyzed databases among the ones mentioned above. Anyway, we could also find attempts of using PIAAC (Arntz et al., 2017; Khrobrer et al, 2015), e-CF and the ICT BOK (Fernandéz-Sanz et al., 2017), especially with the aim of building a job-knowledge bridge. Overall, the unstructured databases are less used, also because the pre-processing phase may not be simple. Focusing on the stakeholders, , we detected two main actors: Institutions and Companies. In particular, the need to recruit the workforce whose characteristics accurately reflect their needs, as well as to identify the lack of internal skills, is increasingly strategic for enterprises. Moreover, in this phase of rapid technological evolution, the Institutions have the need (and objective) of: predicting which will be the main consequences on the labor market; facilitating change through ad hoc policies; acting as an effective bridge between the University offer and the market demand. The reported results are the detection of employment

[6]Found at http://online.onetcenter.org
[7]Found at http://www.oecd.org/skills/piaac/
[8]Found at https://www.cepis.org/index.jsp?p=940n=3016
[9]Found at http://www.ecompetences.eu/it/

drivers of change and of the workers at risk of substitution by machines, mainly obtained through econometrics techniques (Frey et al., 2017; Arntz et al., 2017; Autor et al., 2009; Acemoglu et al., 2010; MacCrory et al., 2014); instead, other results as new skill taxonomies or entity extraction (Boselli et al., 2018; Karakatsanis et al., 2017; Alabdulkareem et al., 2018; Colombo et al., 2019, Jerman et al., 2018; Pejic-Bach et al., 2019), comparison (Thompson et al., 2010) and linking are the output of data mining processes (Mirski et al., 2017; Fernandéz-Sanz et al., 2017; Alfonso-Hermelo et al., 2019, Pryima et al, 2018). The main insights that the map suggests are:

- a frequent attempt of improving the effectiveness of the recruitment process through linking ESCO skills (Mirski et al., 2017; Alfonso-Hermelo et al., 2019; Pryima et al, 2018);
- the econometric approaches for predictive purposes (Frey et al., 2017; Acemoglu et al., 2011) and effective policy design (Alabdulkareem et al., 2018; Autor et al., 2009; MacCrory et al., 2014) are mainly founded on O*NET database;
- there are several attempts to build bridges between universities and market requirements, particularly through text mining applied to ESCO (Fernandéz-Sanz et al., 2017; Colombo et al., 2019) but also to O*NET (Thompson et al., 2010);
- our best knowledge does not contain researches about estimation of industry 4.0 impact on workers, or even analyses performed on job descriptions, that may be due to the fact that the process is made internally and the value obtained does not serve to scientific purposes;
- overall, there seem to be more attempts to respond to the needs of the public sector than the private one.

2.4 Automatic Tools to support Human Resource Management

2.4.1 An overview

More than 50 years ago, the pioneering work carried by the US Bureau of Labor Statistics (BLS) made governments in most developed economies aware of the strong need for assessing national labour market prospects and skills projections (Wilson, 2013). Today, the proactive adaptation which leaders must begin to consider, also embraces the need to manage skill requalification and disruption. From a managerial point of view, the presence of the wrong skills mixes leads enterprises to have bad performances (Grugulis et al., 2009; Lorentz et al., 2013). However, as organizations are complex political entities, determining the right mix of skills for a firm is not an easy task (Abbott, 1993). In light of this, it is increasingly necessary to design tools and methods to spot talents and skill gaps: the insights are useful for several objectives and they help organizations to maximize their profits and to improve their HR management strategies (World Economic Forum, 2016). For an effective application of the technologies, it is strictly necessary to carefully observe company preconditions. The latter have to be properly identified in order to integrate the bases of Industry 4.0 (Crnjac et al., 2017). As we can infer from previous works, there have been many attempts to implement methodologies for managing the technological changes of I4.0. New methods for HR management in the context of I4.0 are clearly rarer, maybe because the impact on the labor market is definitely more complex to be measured and because the phenomenon is relatively new. However, new methods and techniques, such as Data Mining and Analytics, represent a fundamental support to reach the previous goal. In Europe, skill forecast is one of the main issue carried on by the European Centre for the Development of Vocational Training (Cedefop). By using harmonized international data and common approaches for European countries, it provides every year not only quantitative projections of the future trends in employment, but also cross-country comparisons. Other works start from data to predict the future of labor market. Among these, authors can mention the report “Skills Needs Analysis for Industry 4.0 Based on Roadmaps for Smart Systems” (HartmannBovenschulte, 2013) in which, starting from some roadmaps linked to Smart Systems, skills demand is derived and quantitatively explored by constructing scenarios and technology/sector matrices. In the study “Technology usage, expected job sustainability, and perceived job insecurity” (Nam, 2018), data analysis supports the prediction of job insecurity in United States, starting from data collected by telephone survey.

2.4.2 A focus on Text Mining techniques

Some of the emerging tools are based on Text Mining, which represents the semi-automatic process of knowledge extraction from text. The automatic analysis of text is called Natural Language Processing (NLP). The NLP approach usually involves the execution of a software pipeline composed of steps with the aim of extracting information from text. These approaches have shown great success in different fields of research such as mining and summarizing customer reviews (Hu and Liu, 2004), study of social media (Hong et al., 2010) and patent analysis (Chiarello et al., 2018; Lee et al., 2009; Ma et al., 2017; Tseng et al., 2007) and they are now originally applied for detecting skills and job profiles 4.0 (Fareri et al., 2018). In a context of HR support, the Text Mining tools are designed, created and applied to perform both (I) recruiting and (II) firm performance improvement, term that covers many aspects such as market value, growth, employees and customers satisfaction and social and environmental performance (Selvam et al., 2016). The authors try to summarize these two different goals and TM methods used for achieving them in Figure 4.

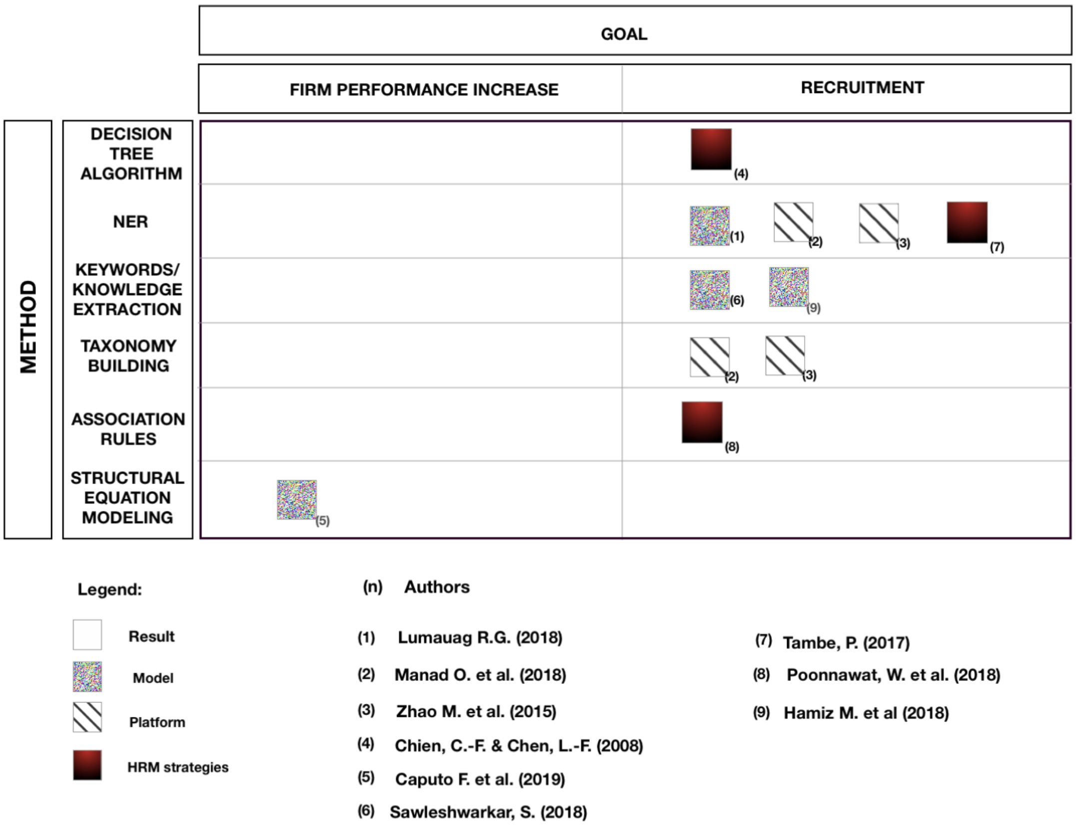


Figure 4: Text Mining Literature Map, representation by authors. The map should be read as follows: the author "n" tries to achieve the goal "k", offering "p" as the main result of his analysis..

For what concerns recruitment, several works (Manad et al., 2018; Sawleshwarkar et al., 2018; Tambe, 2018; Poonnawat et al., 2017; Hamiz, 2017; Lumauag, 2019; Zhao et al., 2015; ChienChen, 2008) try to answer to the ever-growing need of companies to improve and to accelerate the process. To do so, they typically follow the standard workflow of data mining techniques application (i.e. data preprocessing, data analysis and data visualization) and they extract valuable information to support HRM in head-hunting. Instead, some works (Caputo et al., 2019) more generally focus on the improvement of firm performances.

The main methods used in literature in order to pursue the two above-mentioned goals are:

- Decision Tree Algorithm, that is a widely spread algorithm used to classify datasets, following human-readable classification rules (Ben-Haim Tom-Tov, 2010);
- Named Entity Recognition (NER), that is an Information Extraction (IE) technique used to recognize information units (like names of people, organizations, locations) and numeric expressions (like time, date, money and percentages) in unstructured texts (Nadeau Sekine, 2007)
- Keywords/Knowledge Extraction, that are, respectively, (I) the extraction of relevant words of the text in order to summarize it and (II) their formalization in a machine-readable format;
- Taxonomy building, that is the automatic construction of hierarchic classifications from unstructured texts;
- Association Rules, whose identification leads to the discovery of interesting relations between variables in databases (Piatetsky-Shapiro, 1991).

The principal outputs provided by papers are models, platforms and Human Resource Management strategies.

By analysing the map, the main insights authors could mention are:

- NER is frequently used to improve recruitment process;
- text mining methods usually support the development of platforms, that are easy to access, scalable and that allow data gathering and analysis;
- the increase of firm performance seems to be a more marginal objective, probably because more general with respect to the other;
- the goal most of the papers try to reach is the one linked to recruitment. The reason of that could be that the recruitment process is becoming more and more time consuming and not efficient, because of job changes and employee dissatisfaction (Sawleshwarkar, 2018).

2.5 Methodology

In this section, we describe the methodology adopted to assess the impact of Industry 4.0 on Job profiles. As we stated in the introduction, Whirlpool represents an excellent case study for two main reasons: it is a multinational company with an advanced system of skill mapping; the sector to which it belongs (white-goods), is one of the key fields of Italian Economic development (Paris, 2012) and it already began its digitization process, especially thanks to its size and revenue. The designed process is based on Text Mining techniques and a large knowledge base of technologies and methods 4.0 (Chiarello et al., 2018).

The process is graphically summarized in Figure 5. Once the job profiles were pre-processed, we tagged the texts searching for technologies 4.0 and techniques. From these structured data, we then developed a measure able to quantify the readiness 4.0 of each job profile. The final step of the process was to consider the metadata of the job profiles with the aim of understanding if there are differences between 4.0-ready and non-4.0-ready at a skill level. The next sections will explain in greater detail each of the main activities of the process.

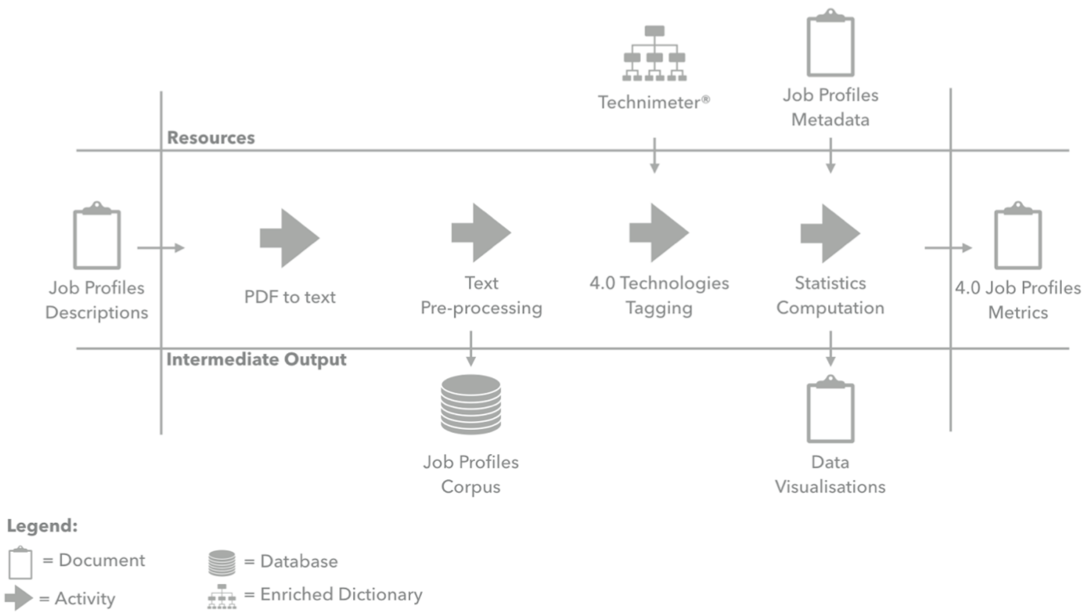


Figure 5: Workflow of the adopted methodology. The central section shows the main process; the upper section shows the external resources used to carry on the analysis; the bottom section shows valuable intermediate outputs.

2.5.1 Documents Import

The first activity to perform in a text analysis pipeline is to collect all the documents that contain useful information for the analysis and then import the set of documents into the computer program. Dealing with Human Resources related documents, this phase is not trivial since usually companies: do not have a structured database of job profiles and the data, if available, are not in a processable format.

The former problem is more a management related problem, thus it is out-of-scope for the present methodology (for this reason the process depicted in Figure 5 starts with a collection of job profiles). On the other hand, the latter can be tackled using various computational techniques.

For the purpose of the present methodology, documents containing job-descriptions are considered in a digital format, and there is no need to read it from an analogical source. From a computer science point of view, text is a human-readable sequence of characters and the words they form that can be encoded into computer-readable formats. There is no standard definition of a text file, though there are several common formats. The most common types of encoding are: standard definition of a text file, though there are several common formats (e.g. txt, doc, html, htm, pdf). In the context of analyzing documents coming from companies, the most typical scenario is the need to transform a document from a pdf format to a plain text format. Even if on the market there are many softwares to carry out this task (e.g. ABBYY FineReader or PDFElement), to ensure the reproducibility of the present work we adopted an open source package developed in the R software (R Core Team, 2018; Wickham H. et al., 2018).

2.5.2 Description of the documents structure

The input data were represented by a set of job descriptions (78) belonging to Whirlpool. The data contained a list of competencies and skills, required to perform a specific role in the organization.

Job Title:	Energy Engineer	**Previous Experience**:	H&S Engineer
Supervisor:	Energy Manager	**Level**:	Leading Self

Job Mission	Everyday Execution	Level
"Lead the implementation of EMEA Remediation and Energies strategies by providing leadership and expertise within EMEA. Improve environmental and energy performances through conservation of natural resources, waste minimization and prevention of pollution."	Business Acumen	Proficient
	Co0pex	N/A
	Change Management	Proficient
	Computer Literacy	Master
	English Language	Proficient
	Problem Solving	Proficient
	Process Excellence	Proficient
	Project Management	Proficient
Main Activity	**Operational**	**Level**
- Support contract and tariff definition - Develop and maintain an updated database with energies data (consumption, costs) - Develop and maintain knowledge of available energies sources and technologies and knowledge of the energy regulatory frame and public incentive programs - Support implementation of E&E strategy - ...	Price&Margin Realization	Basic
	Cost Leadership	Basic
	Customer Quality	Basic
	Innovation	Basic
	Leadership	**Level**
	Leadership	Leading Self
	Functional	**Level**
	WPS	Proficient
	Material Handling	Basic
	Manteinance Expertise	Basic
	...	...

Figure 6: An example of a job description and its structure

Each job profile was reported in a page only, all divided into two columns. In the first column, we found a description of its job mission and its customized main activities (different for each profile); in the second column, we could find its competencies divided in four groups (everyday execution, operational, functional and leadership). Everyday execution are 8 transversal skills related to the deployment of operative activities required by the company processes in day-to-day life; instead, the operational are four and they represent a mix of soft and hard skills which are able to influence both individuals working processes and Company ones. The functional cluster consists of skills, techniques and abilities to deploy operative activities required by that precise role; they vary according to different functions. Finally, leadership skill comprises itself only. The different skills were 40 in total and they were almost the same for every profile, except for the composition of the functional cluster and rarely the operational. Moreover, a level of proficiency was associated with each skill (basic if he/she knows the skill; proficient if he/she applies it, master if he/she teaches it; n/a if he/she does not know/apply/teach). Finally, the level attributed to leadership cluster was binary: leading self for operative roles, leading others for managerial ones.

At the top of the job profiles, the job title, the supervisor, the previous experiences and the level, directly linked to the leadership level described above, could be found (Fig. 6).

Because of the heterogeneity of job missions and main activities, the amount of text varies in relation to the job profile considered. Moreover, the main activity formulation was not easy to manage. In fact, it always starts with a verb that implies several criticality on applying a standard POS-tagger. For these reasons, it seems necessary to define ad hoc rules to manage the problems.

2.5.3 Text Pre-Processing

In the field of Natural Language Processing, text pre-processing means converting documents in a form that is more convenient for the analysis. Most of the task of document analysis in fact relies on first separating (tokenizing) sentences and words, stripping suffixes from the end of words, determining the root of a word or transforming the text using regular expressions.

For the present methodology not all the usual text pre-processing phases (sentence splitting, tokenization, stemming or lemmatization) are required. In particular, since in the next phase we will search for technologies 4.0 in the job profiles description, the only tasks that are required in this phase are Lemmatization and Part of speech tagging. Lemmatization is determining the root of a word. The output allows to find that two words have the same root, despite their surface differences. (For example, the verbs am, are, and is have the shared lemma be; the nouns cat and cats both have the lemma cat.) The Part of speech tagging provides information concerning the morphological role of a word and its morphosyntactic context: for example, if the token is a determiner, the next token is a noun or an adjective with very high confidence.

The output of this phase is a corpus of job profiles that can be taken as input for the technologies tagging phase.

Technology 4.0	**Regular Expression**
actroid	(?i-mx:\b(?:actroid\b))
actuator	(?i-mx:\b(?:(?:actuator—actuators)\b))
adaptable robotics	(?i-mx:\b(?:adaptable(?:\s*robotics\b)))
agricultural robot	(?i-mx:\b(?:agricultural(?:\s*?)robot\w*\b))
android	(?i-mx:\b(?:android\w*\b))
automaton	(?i-mx:\b(?:(?:automaton\w*—automata—automatron)\b))
autonomous robot	(?i-mx:\b(?:autonomous(?:\s*?)robot\w*\b))
biomechatronics	(?i-mx:\b(?:bio(?:\s*?)mechatronics\b))
cobot	(?i-mx:\b(?:(?:cobot—co(?:\s*?)robot—cobots—co(?:\s*?)robots)\b))
drone	(?i-mx:\b(?:(?:drone—drones)\b))

Table 1: An extraction of technologies and their Regular Expressions from research team's taxonomy

2.5.4 Technologies 4.0 Tagging

This phase takes as input the corpus of job profiles described in section 2.5.3 and searches for all the industry 4.0 technologies contained in the text. The word searched and extracted in the Job Descriptions belongs to a taxonomy of Industry 4.0 technologies previously developed (Chiarello et al, 2018). The taxonomy is composed of technologies and methods 4.0 and relations among them. The sources of the system are newspapers, scientific articles and public dictionaries available from academic sources and open databases. For each technology 4.0, the taxonomy contains a regular expression [10] that allows the tool to capture the different orthographic declinations in which a technology could be written. An example of regular expression for the first 10 entries of the taxonomy is shown in Table 1.

The output of this phase is the data structure as shown in Table 2, where for each job profile we have a list of technologies and techniques 4.0. As it is evident from this table not all the 78 job profiles contain at least one technology 4.0, but only 20% is covered. This first result will be discussed in more detail in section 2.6.

2.5.5 Statistics Computation

The authors choose to split job profiles in two groups, assuming that: a job profile could be defined "4.0", if it contains at least one technology 4.0 in the description of its main activities; a job profile could be defined "non-4.0", if it not contains technologies 4.0 in the description of its main activities.

Even if more sophisticated classification of the Job Profile could be chosen (3 or more classes), the approach of binary classification gives us a first high-grain

[10] A regular expression, regex or regexp (sometimes called a rational expression) is a sequence of characters that define a search pattern. Usually this pattern is used by string searching algorithms for "find" or "find and replace" operations on strings, or for input validation

Job Title	Technology 4.0
Capital Management Planner	Tracking
Category Cost Manager	Tracking; Predictive Analytics
Ets Design Engineer	Predictive Analytics
Help Chain	Predictive Analytics
Maintenance Shift Supervisor	Predictive Analytics; Smed
Maintenance Specialist	Predictive Analytics
Manufacturing R&D Manager	Nmbp; Advanced Manufacturing; Smart Manufacturing Systems; Mechatronics; Real-Time Computing; Simulation
Process Technology Assembly / Testing Engineer	Advanced Manufacturing
Process Technology Assembly /Testing Lead Engineer	Advanced Manufacturing
Process Technology Engineer	Advanced Manufacturing
Process Technology Lead Engineer	Advanced Manufacturing
Process Technology Manager	Advanced Manufacturing
Site Maintenance/Toolshop Mgr	Predictive Analytics
Tool Maintenance Supervisor	Smed
Wps Central Pillar Leader	Tracking

Table 2: Technologies 4.0 related to each job profile

division between the job profiles. In this way we will have a high recall to find signal of digitalisation in job profiles, even if highly digitized professions (that have many digital competencies) and poorly digitised ones (that uses only one 4.0 technology in their job) are considered the same. As stated in section 2.5, for each job profile, we have not only the associated skills, but the level of proficiency required by each profile too. The levels of proficiency are: n/a (level 0), basic (level 1), proficient (level 2) and master (level 3). We are interested in looking at differences between the job profiles 4.0 and the job profiles non-4.0 for what concerns the different types of skills. In particular we look for any differences between the two groups of job profiles in terms of: everyday, operational, leadership and functional skills (aggregate) and specific skills inside each of the aforementioned skills groups.

The statistics to address these two levels of analysis are described in sections Skills Groups Statistic and Specific Skills Statistics.

Skills Groups Statistics As described in section 2.5.3 each skill level is measured on a scale from 0 to 3. Thus, we computed the mean ms and the standard deviation s for 6 group of skills s: everyday execution, operational and functional skills for profiles 4.0 and the non-4.0 ones. Then, to have a measure of the dispersion around the mean of the level of each skill s, we computed two measures:

Smins=ms - s Smaxs=ms + s

Finally, since Smins could be negative and since Smaxs could be greater than 3 (the maximum level assigned to the skill) we rounded the minimum of each Smins to 0 if Smins¡0 and of each Smaxs to 3 if Smaxs¿3. This gives a more standard representation of the level of the skills.

Specific Skills Statistics The statistics for the specific skills are computed in a similar way with respect to the group. In fact, we computed the mean ms and the standard deviation s for 80 group of skills s: each of the 40 skills (described in section 2.5.3) for 4.0 and the non-4.0 profiles. Then we computed Smins and Smaxsas described in section 2.5.5.

2.6 Results

2.6.1 Job profiles and technologies 4.0 relations

The authors started from the assumption that if a job description contains a technology, the correspondent job profile should manage it, somehow. Thus, for each technology we also have the correspondent skill, whose purpose (to know, to handle, to use, to control, etc. . .) changes according to the specific job title[11]. The Figure 7 represents the first output of the analysis performed and it shows the relationships among job profiles and different technologies belonging to our knowledge base. The connection is visually represented only if the job description contains one of the technologies.

[11] The job title is the term that identifies the position held by the employee

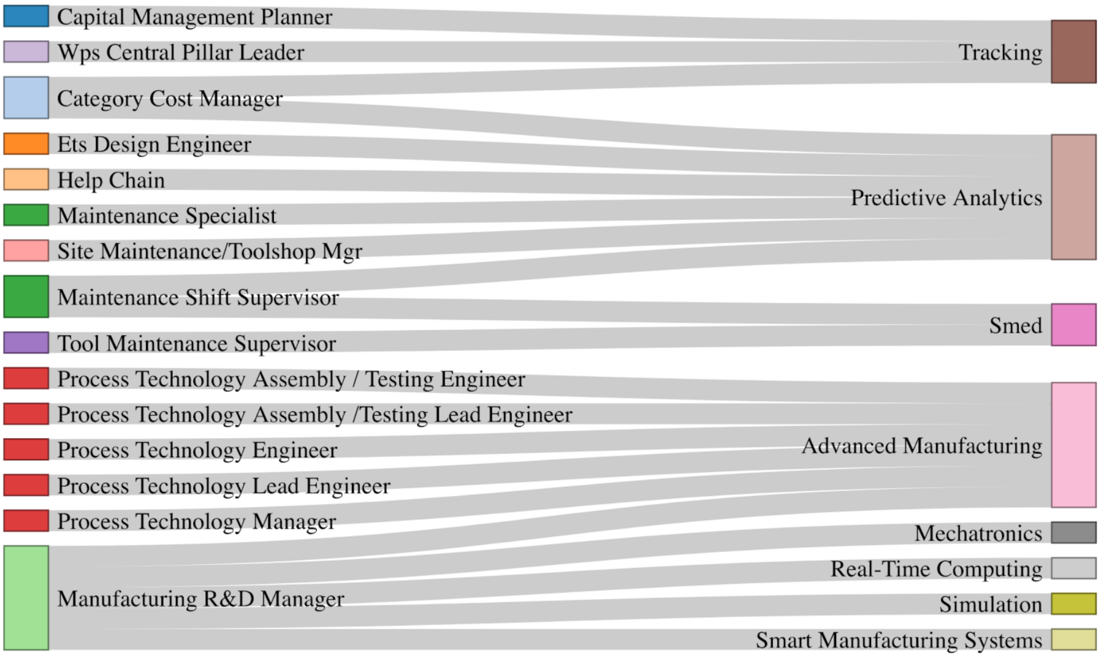


Figure 7: Graphic representation of the technologies 4.0 related to each job profile

As mentioned in section 2.5, only 15 job profiles (20% of the total) contain at least one technology 4.0. The queues are interesting: all job profiles that perform Process Technology are linked to the technology Advanced Manufacturing. Vice versa, the mechatronics and simulation technologies, which are in the technologies queue, are all linked to a single profile: the Manufacturing RD Manager. Another peculiarity is represented by the strong connection between profiles and the technology Predictive Analytics. Moreover, the image reveals that the analytics tool is mostly used in accounting and maintenance. Another relevant evidence emerges from table 3, where we count the percentage of 4.0 and non 4.0 job profiles for different leading levels of the hierarchical structure of the company. From this table is evident the bottom-up approach to competences adoption in the context of Industry 4.0. We have in fact an increasing trend in the percentage of job profiles that master I.40 technologies going from high level managers to operational profiles. In the complex and fast-paced world of manufacturing 4.0, there is rarely ever one solution to a problem. For this reason, table 3 gives and evidence that in some contexts I4.0 is taking an approach that is similar to lean manufacturing techniques in the 20th century (Shan et al., 2003), an undeniable use-case for the success of democratized knowledge and a bottom-up way of working.

Leading Level	**Non 4.0**	**4.0**
Leading a Function	100%	0%
Leading Others	82%	18%
Leading Self	71%	29%

Table 3: Percentages of 4.0 and non 4.0 job profiles for different leading levels (leading a function, leading others and leading self) of the hierarchical structure of the company.

2.6.2 Job profiles 4.0 and non-4.0

The following graphs show the differences between profiles 4.0 and non-4.0 comparing their skills. The latter belong to the three groups (everyday execution, operational and functional). A profile is considered 4.0 if it contains at least one technology 4.0 in its job description; otherwise, not-4.0.

Everyday Execution Skills Figure 8 represents the mean, the maximum and minimum level for Everyday Execution Skills, for profiles 4.0 and not-4.0.

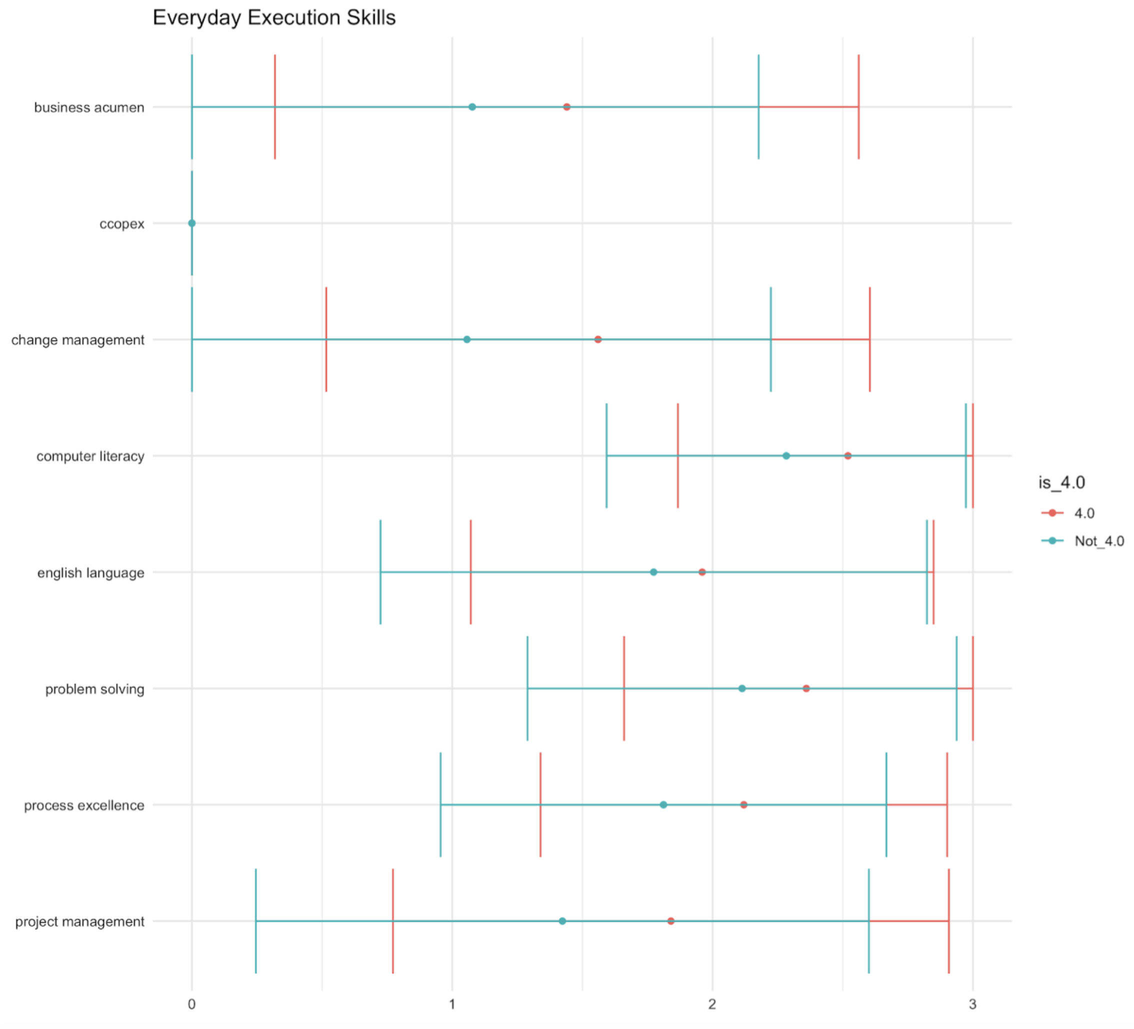


Figure 8: Representation of the mean and range distribution of everyday execution skill for job profiles 4.0 and non-4.0

All the skills belonging to the Everyday Execution cluster are transversal and many of them are methodological (e.g. Project Management and Change Management). Moreover, particularly relevant are Business Acumen, Project Management and Change Management, in which the gap between profiles 4.0 and non-4.0 is substantial. Furthermore, for the majority of skills the range distribution is wider for non-4.0 profiles. The skills for which the profiles 4.0 and the non-4.0 ones have fewer differences are the English Language and the Computer Literacy. Finally, Problem Solving has a high value both for profiles 4.0 and non-4.0. Therefore, also Problem Solving seems to be taken for granted within the company. However, it would be necessary to understand to which extent it is considered in its methodological completeness.

Operational Skills As previously mentioned, the cluster of operational skills consists of two hard skills (Cost Leadership and PriceMargin Realization) and three transversal skills (Continuous Improvement, Customer Quality and Innovation). Figure 9 represents the mean and the range distribution of Operational Skills, for profiles 4.0 and not-4.0.

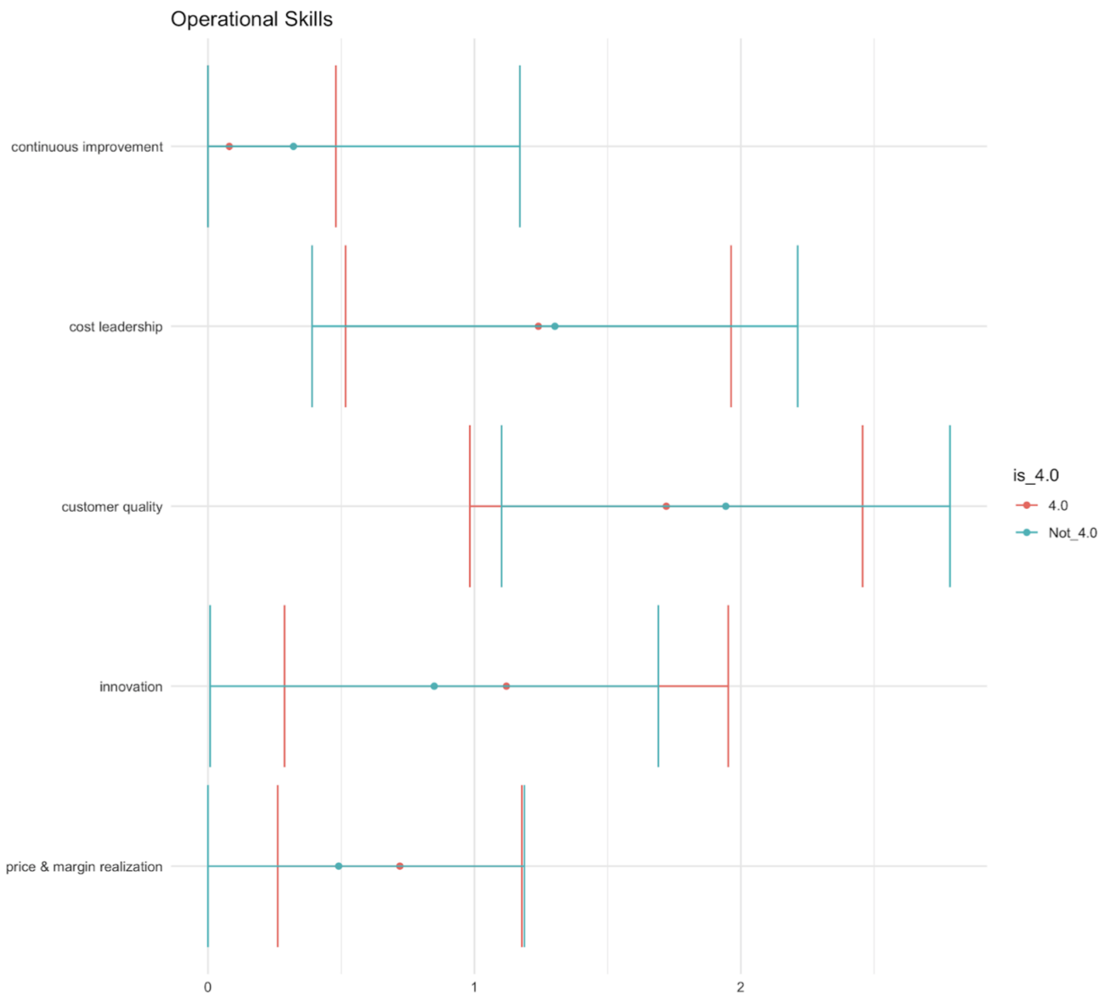


Figure 9: Representation of the mean and range distribution of operational skill for job profiles 4.0 and non-4.0

While for all the skills belonging to Everyday Execution cluster, profiles 4.0 had a higher mean score than non-4.0 ones, for the skills belonging to the Operational cluster the data provide different insights: the job profiles non-4.0 have a higher score than the 4.0 ones. In particular, the Customer Quality and Continuous Improvement are averagely mastered at a higher level by job profiles non-4.0. Instead, skills such as Innovation and Price and Margin Realization remain the prerogative of professional profiles 4.0. Therefore, the job profiles 4.0 are stronger on Innovation and Price Margin Realization; it is very consistent with their role as "managers and users" of emerging technologies. A strange anomaly is represented by the level of the Co0pex[12] skill, which is almost zero, as it only appears once within the job profile descriptions.

Functional Skills The cluster of Functional Skills contains both hard (as Logistic Expertise or Material Flow Design) and transversal skills (as Industrial Relations and PDCA). Figure 10 represents the mean and the range distribution of functional skills, for profiles 4.0 and not-4.0.

[12] ability to drive out waste and improve the quality, cost and time performance of business processes

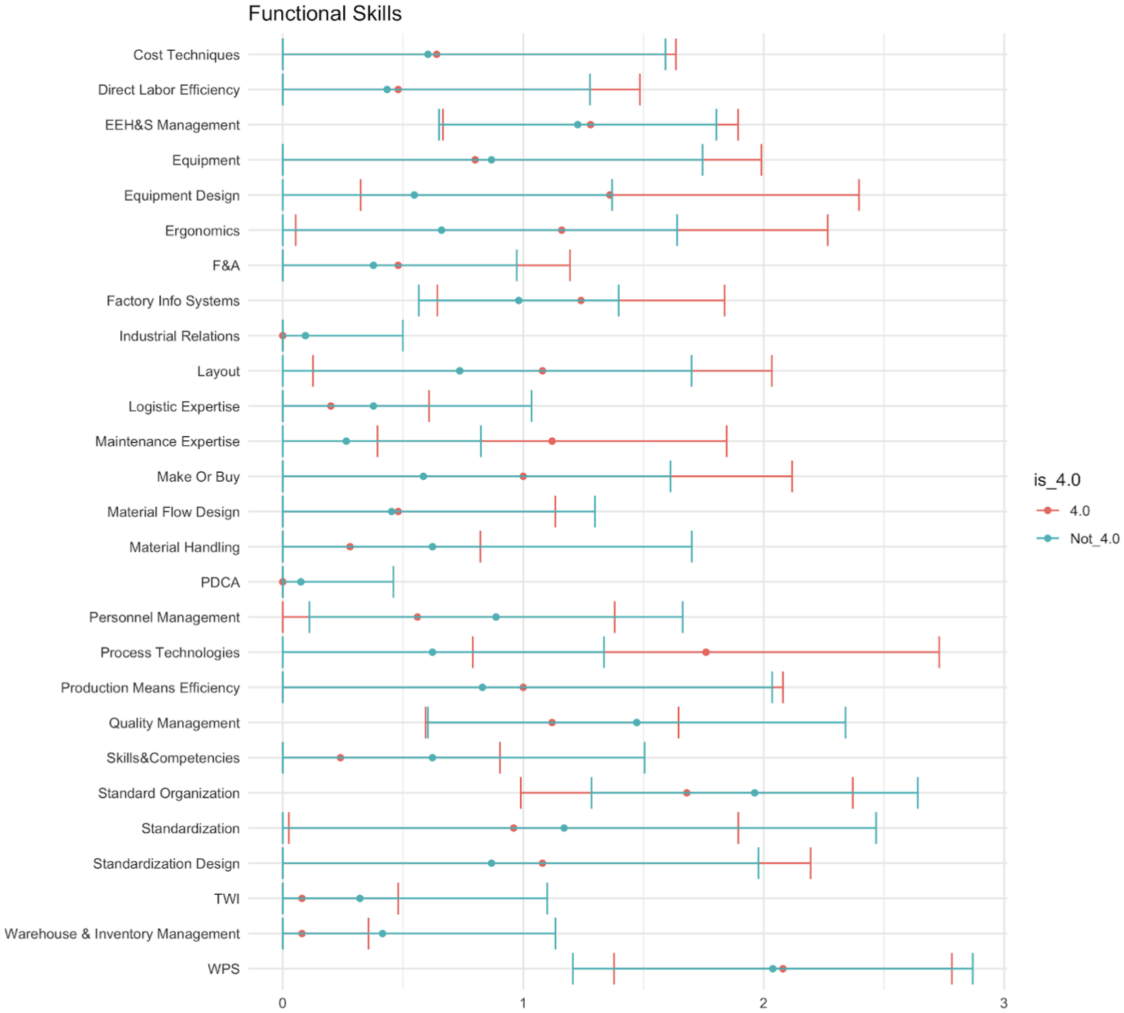


Figure 10: Representation of the mean and range distribution of functional skills for job profiles 4.0 and non-4.0

Industrial Relations and PDCA are managed at a higher level by job profiles non-4.0. This seems consistent for Industrial Relation skills, which rely on both the ability to manage relations inside and outside the organization and the ability to apply the tools and methods to analyze and define spaces and relations. For this reason, this competence is only partially linked to technologies and it also depends on personal characteristics. Instead, the Deming Cycle PDCA (Plan-Do-Check-Act) should be managed at the same level by the profiles 4.0 and non-4.0, as it is a method used for all business processes and that can be largely supported by new technologies. The TWI (Training Within Industry), is a dynamic program of hands-on learning and practice, teaching essential skills for people who lead the work of others in the enterprise. The latter helps people to be able to contribute to the training process: the profiles non-4.0 have a higher mean than the 4.0 ones, maybe because the ability to train others is mostly independent from technologies 4.0. Job profiles 4.0 present a higher average level of skills such as Process Technologies, Equipment Design, Factory Info Systems, Ergonomics and Maintenance Expertise. For example, the average level of the skill Maintenance Expertise for profiles 4.0 is higher than its distribution peak level for profiles non-4.0. In general, all these skills require the knowledge of technologies 4.0 in order to be better applied. It seems consistent, because they are strictly related to the technologies: the most fitting example is the skill Factory Info Systems, or the ability to manage the hardware and software needed to collect, process and share data inside and outside the organization. It is also interesting to notice that skills such as Quality Management and Warehouse Inventory Management have a higher average level for profiles non-4.0. Actually, the processes concerning the quality and the warehouse management present many characteristics that make them particularly suitable for technologies 4.0 introduction. The previous results deserve reflection and they are an important opportunity for improvement.

2.6.3 Results Validation

Our classification of job profiles (4.0 and non-4.0) was validated through the information available on O*NET (The Occupational Information Network), which is an open-source database developed for the U.S Department of Labor, made of 974 occupations from Standard Occupational Classification (SOC). O*NET categorizes a job as bright outlook if it is expected to grow rapidly in the next several years or will have a large number of job openings. Every bright outlook occupation meets at least one of the following criteria: its employment will increase of 7% (or more) over the period 2018-2028 or it is expected to have more than 100,000 job openings in the same period[13]. Furthermore, O*NET tracks if an occupation owns hot technology skills, defined as technology requirements frequently included in employer job postings. For each Whirlpool Job Title 4.0, we searched for its correspondent occupation on O*NET, extracting the number of hot technology skills owned and tracking if it was labelled as bright outlook or not. We did the same for 8 random profiles categorized as non-4.0, to double check our outcomes. The results are reported in Table 4.

First of all, it is evident that 19 on 23 Whirlpool job profiles has an exact match on O*NET, and the others are present in the database as alternative labels of the job profile. This is an evidence of the fact that the Whirlpool Job description database is aligned with O*NET and thus the database is a good source for validating our

[13]Bureau Labor Market statistics, 2018-2028 employment projections (source: https://www.bls.gov/emp/)

Whirlpool Job Title	Our Classification	O*NET Job Title	Matching Type	O*NET Classification	Number of Hot Technology Skills
Capital Management Planner	4.0	Financial Manager	Alternative Label	Bright Outlook	35
Category Cost Manager	4.0	Cost Estimators	Alternative Label	Bright Outlook	18
Ets Design Engineer	4.0	Electrical Engineering Technologists	Alternative Label	\	15
Help Chain	4.0	Helpers-Installation, Maintenance, and Repair Workers	Alternative Label	Bright Outlook	8
Maintenance Shift Supervisor	4.0	Shift Maintenance Supervisor	Exact	Bright Outlook	17
Maintenance Specialist	4.0	Maintenance Support Specialist	Exact	Bright Outlook	17
Manufacturing R&D Manager	4.0	Manufacturing Engineering Manager	Alternative Label	\	10
Process Technology Assembly / Testing Engineer	4.0	Test Engineer	Exact	Bright Outlook	27
Process Technology Assembly / Testing Lead Engineer	4.0	Test Engineer	Exact	Bright Outlook	27
Process Technology Engineer	4.0	Industrial Engineer	Alternative Label	Bright Outlook	28
Process Technology Lead Engineer	4.0	Industrial Engineer	Alternative Label	Bright Outlook	28
Process Technology Manager	4.0	Manufacturing Operation Manager	Alternative Label	Bright Outlook	59
Site Maintenance/ Toolshop Mgr	4.0	Maintenance Manager	Exact	Bright Outlook	17
Tool Maintenance Supervisor	4.0	Maintenance Supervisor	Exact	Bright Outlook	17
Wps Central Pillar Leader	4.0	Plant Manager	Alternative Label	Bright Outlook	59
Audit Manager	non-4.0	Audit Manager	Exact	Bright Outlook	30
Receiving Office Operator	non-4.0	Receiving Office Operator	Exact	\	8
Laboratory Operator	non-4.0	Laboratory Operator	Exact	\	4
Quality Inspecting Operator	non-4.0	Quality Inspector	Exact	\	19
Value Stream Manager	non-4.0	Value Stream Manager	Exact	\	9
Production Assistant	non-4.0	Production Assistant	Exact	\	7
Material Planner	non-4.0	Material Planner	Exact	\	24
Inbound Team Leader	non-4.0	Inbound Team Leader	Exact	\	8

Table 4: Correspondence between Whirlpool Job profiles and O*NET (Label, Classification and Hot Technology Skills owned). A job profile in O*NET is labelled as "bright outlook", if it is expected to grow rapidly in the next several years while a hot technology skill is a hard skill increasingly required by the market. The column matching type shows if Whirlpool Job title matched O*NET job title or if it matched an alternative label.

tool. The validation process returned 21 to 23 matches (91,3%). Generally, the profiles labelled with bright outlook are characterized by a consistent number of hot technology skills (higher than 15), that was in line with our expectations. The presence of a job profile wrongly labelled as non-4.0 indicates a technology that is currently missing on Whirlpool database. This underlines the necessity to update Whirlpool Job descriptions, that should be aligned with Competence Frameworks and the ever-changing labor market language. For what concerns the false positive Manufacturing RD Manager, it represents an interesting exception because it remarks the importance of analyzing a hard skill considering also the verb that precedes the technology. In that case, the job profile knows but not use the technology, so it is not tracked as relevant on O*NET.

2.7 Discussion, Implication and Conclusions

2.7.1 Discussion

The article showed that the profiles 4.0-ready seem to be more adequate to carry out the operative activities required by the Company processes in day-to-day life through the relevant support of their soft skills. The previous gap is probably due to the key role that Change Management and Project Management will cover in a context of renewal. Moreover, job profiles 4.0 are stronger on all transversal skills; we can assume that, for managing (and facing) the introduction of new technologies, there should be an important soft component (Chui, 2016). Furthermore, the so-called Soft Skills are increasingly central in the current digital era. The demonstration of how much job computerization will affect those who perform routine-based tasks (Frey et al., 2017) underlines the critical role of transversal skills, as a key for withstanding the (likely) penalizing effects on employment. For the majority of skills the range distribution for everyday execution skills is wider for non-4.0 profiles. A possible explanation to the previous gap is represented by the quite relevant differences in proficiency level among 4.0 and non-4.0 profiles. Instead, no differences are detected for English Language and Computer Literacy. This evidence confirms that these two skills are now almost essential within companies and always requested. (Azmi, 2018) Fifteen profiles are impacted by technologies 4.0, which demonstrates the increasing interest about the topic and makes it possible to have some insights about the impact these technologies are having on other skills. It is evident from Whirlpool data that the profiles 4.0-ready are strong in Innovation and Price Realization, probably because a huge return of capital is required to invest on new expensive equipment. Among the essential skills 4.0, the above mentioned ones cannot be lacking. Another peculiarity is represented by the strong profile connections with the technology Predictive analytics, the most frequent technology 4.0 detectable. This is due both to the fact that this is a general purpose technology and that skills needed to manage data and make predictions based on their analysis (especially to perform predictive maintenance) are the priority for actually beginning the change (Roy, 2016; Colegrove, 2017). The RD manager is the worker that is most impacted by new technologies because he/she is the one that handles innovation in Company; we can assume that RD manager is the one who knows the emergent technologies (even if does not use them) thus plays an essential role to manage the change inside the company.

2.7.2 Implications

Managerial Implications The proposed tools could be used by HR managers (main beneficiaries of this work) to collect information about key professional figures 4.0 and skill 4.0 gaps of the company. This gives to the manager a data driven map of the competences 4.0 in the company and makes it evident which is the impact that technologies 4.0 are having and will have in the near future. Considering this map, different actions can be taken in order to start a data-driven improvement of the HR management process.

First of all, the tool can help automatize the revision and the (continuous) integration of the job profiles. Spotting missing technologies can help HR managers in order to make periodical revisions of the job descriptions they make, and renovate them considering which new technology must become a skill of the workers. Given the speed at which technologies are changing, HR managers will need to implement a process in order to make this periodical revision more efficient. Automatisation is clearly a good solution in this direction.

For what concerns learning programs, they can be customized according to the skills already owned by the job profile. In this sense, the tool can help to automate the first steps of the processes of reskilling and upskilling (learning objectives and learning path identification). In fact, the knowledge base of technologies and methods 4.0 (Chiarello et al., 2018) could support the design of training courses since it contains not only technologies and methods, but also the relation among them. For this reason, it is possible to highlight the best way to acquire a new skill, starting from the ones possessed and following the paths between them. This is handy for HR managers because the ever changing technological environment makes it hard to always have a clear view on which are the technologies 4.0 relevant for a company. The HR manager cannot always be updated on the technological changes that are evolving in the many subfields of industry 4.0 (e.g. cloud computing, big data, additive manufacturing, cyber-security). The tool can thus be considered a knowledge enhancement tool for spotting important digital skills and help HR managers in defining the best path in order to collect these competences.

Finally, the results of the analysis could also underline the need for a bigger set of new and related competences. This sometimes can be considered as the need to hire one or more figures. For reasons that are similar to the ones described before (ever changing digital environments and being updated on the technological changes), it can be really hard for HR managers to write and effective job vacancy. From this perspective, this work contributes to the improvement of the recruitment process, highlighting companies' 4.0 needs and allowing the building of more suitable vacancies.

Academic Implications The proposed paper gives academic contributions in two main fields: Human Research Management and Data Science. The work in fact gives a tool that can make it possible for HR scholars to have a comprehension of the impact of Industry 4.0 on a specific set of documents. These documents can be both internal of the company (e.g. job profiles, internal standards or exhaust data) or external (e.g. social networks, scientific papers, patents) opening to a wide range of research questions to which scholars can answer using the proposed tool. Furthermore, to the reach of our knowledge this research is the first that explores industry 4.0 impact on skills and job profiles by analyzing job descriptions based on a case study. The literature lacks in the analysis of the impact of paradigm 4.0 on workers, in particular by using job descriptions as data source (since this information is usually kept private by the companies). The proposed application could be the pilot for mapping the job descriptions belonging to other companies of different sectors and of different dimensions, in order to have a detailed and shared framework. For what concerns scholars working in the data science field, the proposed work gives evidence that state-of-the art (or even simpler) data science tools can be effective for proposing solutions in the field of HR. This is of great interest for people working in this domain, since it can open to a cross field of study. In particular, the paper makes it evident that text mining techniques, usually performed on structured and generic databases such as O*NET and ESCO, can also be applied on unstructured and domain-specific sources if supported by external knowledge bases. The use of data science in the context of HR studies, is in fact opening to the development methodologies that are scalable and replicable, overcoming the typical issues of replicability in the Social Sciences.

2.7.3 Conclusions

The pace of change is always faster, and the need for new skills and jobs to manage industry 4.0 impact is increasingly emerging in the digital economy. The International competences databases (such as ESCO and O*NET) and the national/regional ones have to be always updated, in relation to the changes driven by the fast Technological Tsunami and this is not a trivial task. The problem is not only related to the public sector, but most of all to large and medium sized companies that have to update both the inventories of skills and job profiles. However, the state of the art focuses on developing models to assess digital maturity at a technological level: to the reach of our knowledge, no frameworks for estimating the impact tha industry 4.0 is having on human capital had been developed in the literature. This article tackles the criticality from a technical point of view, by proposing a quantitative data-driven methodology tested on a real case-study. The case study was carried out in an Italian plant of Whirlpool, a multinational company with a system of skill mapping that is particularly advanced in both the inventory level and job profiles. The main reason why the authors chose Whirlpool is that it represents one of the most representative firms belonging to white-goods sector, one of the key fields of Italian Economic development (Paris, 2012); hence, its size and revenue allow the beginning of its digitization process. The performed analyses are based on data science techniques, which make the process cost-effective (especially in execution times), repeatable, and extensible to other contexts. For this reason, having led the case study on Whirlpool allowed us to obtain results that are (I) significant because linked to a company in a traditional sector included in a process of digitization and (II) generalizable to other multinational companies. The technology 4.0 detection on worker's main activities, enabled their clusterization in two key groups: profile 4.0-ready and non-4.0-ready. The data mining process also allowed the evaluation of the differences between 4.0-ready and non-4.0-ready profiles skills. While the average level of the vertical skills such as operational and functional turns out to be quite the same for both 4.0 and non-4.0 profiles, the transversal skills definitely resulted more relevant for the profiles 4.0. Thus, the analysis output confirmed the increasing importance of the horizontal skills (such as soft skills) in the digital era. Furthermore, the developed system has allowed to find small discrepancies within the competence structure and it lets the team redefine a limited number of new skills 4.0. Therefore, it proved to be easily replicable in other contexts, especially where structured systems for managing skills are already implemented.

Limitations and Future Developments In the present section we list the limitations that the authors have identified with respect to the present work. As can be seen, the main identified limitations are related to the algorithm, and can thus be seen as future developments in order to improve the proposed method. First of all, the authors chose to classify a job as "4-0 ready" if it contains at least one technology 4.0. This is an intentional stretch to divide profiles in two distinct classes that brings simplicity and clarity to the proposed process, and have a high recall in terms of 4.0 impacts (we consider and impact even if it is minum). Anyway, this decision can bring to false positives (job profiles that are classified as impacted by industry 4.0 which are actually not) and this approach needs to be refined in the future. In order to do that we will need to integrate information coming from other sources in order to have metrics also on the technologies and not only on the job profile. The NER system will be improved and its extraction capability enlarged in order to have more context with respect to technology. In particular it can be interesting to understand the level at which a job profile has a relation with a certain technology (i.e. know, use, manage, etc...). This can be done by making it possible for our system to detect also these clues around the extrafted technology. Our validation process consisted of comparing our results with the information available in O*NET. We chose to consider relevant the classification as bright outlook and the number of hot technologies owned. Even if O*NET represents one of the most reliable sources on skills and job profiles, the selection of the previous two variables is still subjective. Moreover, the validation process of the results is not embedded nor automatized. In the future works, we will refine our rationale and we will automatize the step performed on O*NET. For what concerns the scope of the analyzed document, it has to be considered that even if the results shown in the present work are quantitatively supported by data, they cannot lead to a more general conclusion about the changes that Industry 4.0 is having on skills and job profiles but it can only give some qualitative insight. To solve these problems, in the future we will work on other kinds of documents, less context specific, such as job vacancies, curricula and scientific papers. To work on these domain the tool will need a relatively small redesign, while it has to be expected an increase in the number of analyzed documented and thus an increased level of statistical relevance. Finally, since the output of the tool is a static representation of the impact 4.0 on human capital, the next step of the research will be understanding how to extract the interconnections among job profiles and skills (both 4.0 and non-4.0 ready). Thus, we would try to define how the information extracted by the tool can have an improvement on other human resources related processes, i.e. up-skilling and reskilling. Finding the interconnections among skills and profiles, it would be possible both to design training courses following a data-driven learning path and, where appropriate, choose internal resources to conduct them, minimizing the learning effort.

2.8 Appendix 1: Values of the mean and range distribution of skills for job profiles 4.0 and non-4.0

Everyday Execution Skill	Mean Skill 4.0	Min Skill 4.0	Max Skill 4.0	Mean Skill NOT 4.0	Min Skill NOT 4.0	Max Skill NOT 4.0
business acumen	1,29	0,14	2,43	1,16	0,05	2,27
ccopex	0	0	0	0	0	0
change management	1,38	0,36	2,4	1,16	0	2,35
computer literacy	2,43	1,75	3	2,33	1,64	3
english language	1,76	0,93	2,59	1,86	0,8	2,92
problem solving	2,24	1,54	2,94	2,18	1,35	3
process excellence	1,95	1,21	2,69	1,89	1,01	2,77
project management	1,62	0,6	2,64	1,54	0,33	2,74

Table 5: Values of the mean and range distribution of everyday execution skill for job profiles 4.0 and non-4.0

Operational Skill	Mean Skill 4.0	Min Skill 4.0	Max Skill 4.0	Mean Skill NOT 4.0	Min Skill NOT 4.0	Max Skill NOT 4.0
Continuous improvement	0,1	0	0,53	0,3	0	1,12
Cost leadership	1,14	0,42	1,87	1,33	0,44	2,23
Customer quality	1,67	0,94	2,4	1,95	1,11	2,78
Innovation	1,05	0,24	1,85	0,89	0,04	1,75
Price & margin realization	0,67	0,18	1,15	0,53	0	1,21

Table 6: Values of the mean and range distribution of operational skill for job profiles 4.0 and non-4.0

Functional Skill	**Mean Skill 4.0**	**Min Skill 4.0**	**Max Skill 4.0**	**Mean Skill NOT 4.0**	**Min Skill NOT 4.0**	**Max Skill NOT 4.0**
Cost Techniques	0,62	0	1,54	0,61	0	1,63
Direct Labor Efficiency	0,48	0	1,51	0,44	0	1,28
EEH&S Management	1,1	0,79	1,4	1,3	0,64	1,95
Equipment	0,81	0	2,06	0,86	0	1,73
Equipment Design	1,33	0,22	2,44	0,61	0	1,45
Ergonomics	0,9	0	1,9	0,79	0	1,85
F&A	0,38	0	1,12	0,42	0	1,02
Factory Info Systems	1,29	0,64	1,93	0,98	0,58	1,38
Industrial Relations	0	0	0	0,09	0	0,48
Layout	0,9	0	1,85	0,82	0	1,81
Logistic Expertise	0,19	0	0,59	0,37	0	1,01
Maintenance Expertise	1,14	0,35	1,94	0,32	0	0,89
Make Or Buy	1	0	2,1	0,61	0	1,66
Material Flow Design	0,52	0	1,2	0,44	0	1,26
Material Handling	0,1	0	0,4	0,67	0	1,72
PDCA	0	0	0	0,07	0	0,44
Personnel Management	0,43	0	1,17	0,91	0,13	1,7
Process Technologies	1,81	0,83	2,79	0,68	0	1,44
Production Means Efficiency	1,05	0,02	2,07	0,82	0	2,04
Quality Management	1,05	0,55	1,55	1,47	0,63	2,32
Skills&Competencies	0,29	0	1	0,58	0	1,44
Standard Organization	1,67	1,01	2,32	1,95	1,26	2,64
Standardization	1,14	0,23	2,05	1,09	0	2,37
Standardization Design	1	0	2,18	0,91	0	2
TWI	0,1	0	0,53	0,3	0	1,05
Warehouse & Inventory Management	0,05	0	0,27	0,4	0	1,11
WPS	2,2	1,28	2,62	2,09	1,26	2,92

Table 7: Values of the mean and range distribution of functional skill for job profiles 4.0 and non-4.0

2.9 References

Abbott, A., (1993). *The sociology of work and occupations.* Annu. Rev. Sociol. 19 (August), 187–209.

Acemoglu, D., Autor, D., (2010). *Skills, tasks and technologies: implications for employment and earnings.* doi:10.3386/w16082

Alabdulkareem, A., Frank, M. R., Sun, L., Alshebli, B., Hidalgo, C., Rahwan, I., (2018). *Unpacking the polarization of workplace skills.* Science Advances,4(7). doi:10.1126/sciadv.aao6030

Alfonso-Hermelo, D., Langlais, P., Bourg, L., (2019). *Automatically learning a human-resource ontology from professional social-network data.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11489 LNAI, pp. 132-145. DOI: 10.1007/978-3-030-18305-9_11

Arnold, C., Kiel, D., Voigt, K.I., (2016). *How the industrial internet of things changes business models in different manufacturing industries.* Int. J. Innov. Manag. 20 (8).

Arntz, M., Gregory, T., Zierahn, U., (2017). *Revisiting the risk of automation.* Economics Letters,159, 157-160. doi:10.1016/j.econlet.2017.07.001

Attardi G., (2006). *Experiments with a multilanguage non-projective dependency parser.* In Proceedings of the Tenth Conference on Computational Natural Language Learning (pp. 166-170). Association for Computational Linguistics.

Autor, D., Dorn, D., (2009). *The growth of low skill service jobs and the polarization of the U.S. Labor Market.* doi:10.3386/w15150

Azmi, A.N., Kamin, Y., Noordin, M.K., Md. Nasir, A.N. (2018) *Towards industrial revolution 4.0: Employers' expectations on fresh engineering graduates.* International Journal of Engineering and Technology(UAE), 7 (4), pp. 267-272. DOI: 10.14419/ijet.v7i4.28.22593

Bakule, M., Czesana, V., Havlickova, V. (2016). *Developing skills foresights, scenarios and forecasts: guide to anticipating and matching skills and jobs.*

Barney, J. B., (1991). *Firm resources and sustained competitive advantage.* Journal of Management, 17: 99-120.

Bauer, C., Viola, K., Strauss, C., (2011). *Management skills for artists: 'learning by doing'?* Int. J. Cult. Policy 17 (5), 626–644.

Becker, B., Gerhart, B., (1996). *The impact of human resource management on organizational performance: Progress and prospects.* Academy of Management Journal, 39: 779-801.

Ben-Haim Y., Tom-Tov E., (2010). *A streaming parallel decision tree algorithm.*

Boselli, R., Cesarini, M., Mercorio, F., Mezzanzanica, M., (2018). *Classifying online job advertisements through machine learning.* Future Generation Computer Systems,86, 319-328. doi:10.1016/j.future.2018.03.035

Bridgstock, R., (2011). *Skills for creative industries graduate success.* Educt. Training 53 (1), 9–26.

Cacciolatti, L., et al., (2017). *Clashing institutional interests in skills between government and industry: An analysis of demand for technical and soft skills of Graduates in the UK.* Technological Forecasting and Social Change

Caputo, F., Cillo, V., Candelo, E., Liu, Y., (2019). *Innovating through digital revolution: The role of soft skills and Big Data in increasing firm performance.* Management Decision. ISSN 0025-1747

Caruso L., (2017). *Digital innovation and the fourth industrial revolution: Epochal social changes?* Ai Society, 33(3), 379-392. doi:10.1007/s00146-017-0736-1

Cedefop, (2016). *Italy: Mismatch priority occupations.* https://skillspanorama.cedefop.europa.eu/en/analytical_highlights/italy-mismatch-priority-occupations

Chen, H., (2017). *Theoretical foundations for cyber-physical systems: a literature review.* J. Ind. Integr. Manag. 2 (03), 1750013.

Chiarello F., Trivelli L., Bonaccorsi A., Fantoni G., (2018). *Extracting and mapping industry 4.0 technologies using wikipedia.* Computers in Industry, 100, 244-257.

Chiarello, F., Cimino, A., Fantoni, G., Dell'Orletta, F. (2018). *Automatic users extraction from patents.* World Patent Information, 54, 28-38. doi:10.1016/j.wpi.2018.07.006

Chien, C.-F. Chen, L.-F.,(2008). *Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry.*

Chryssolouris, G., Mavrikios, D., Mourtzis, D., (2013). *Manufacturing systems: skills competencies for the future.* Procedia CIRP 7, 17–24.

Chui M., Manyika L., Miremadi M., (2016). *Where machines could replace humans and where they can't (yet)*, McKinsey Quarterly, (3), pp. 58-69, https://www.mckinsey.it/file/6255/download?token=eWPahDZw

Colegrove, L., Peters, J., (2017). *Big data analytics skills - A revolution lacking in revolutionaries.* Sustainable Engineering Forum 2017 - Topical Conference at the 2017 AIChE Spring Meeting and 13th Global Congress on Process Safety, pp. 2-10.

Colombo, E., Mercorio, F., Mezzanzanica, M., (2019). *AI meets labor market: Exploring the link between automation and skills.* Information Economics and Policy, 47, pp. 27-37. DOI: 10.1016/j.infoecopol.2019.05.003

Cooper, R., Tang, T., (2010). *The attributes for career success in the mass communication industries: a comparison of current and aspiring professionals.* J. Mass Commun. Educ. 65 (1), 40–55 (Mar.).

Crnjac M., Veža I., Banduka N., (2017). *From concept to the introduction of Industry 4.0.* International Journal of Industrial Engineering and Management, 8, 21-30.

Dengler K., Matthes B., (2018). *The impacts of digital transformation on the labour market: Substitution potentials of occupations in Germany.* Technological Forecasting and Social Change, Volume 137, Pages 304-316, ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2018.09.024.

Dobrunz, J., Schöppner, H., Wolfram, A., (2006). *Anforderungen an den Mediennachwuchs* — die media.net employability-studie. media.net berlinbrandenburg e.V, Berlin.

European Centre for the Development of Vocational Training (Cedefop). (2019). *Online job vacancies and skills analysis: a Cedefop pan-European approach.*

European Commission, (2013). *ESCO - European Classification of Skills/Competences, Qualifications and Occupations.*

Fareri, S., Chiarello, F., Coli, E., Dente, G., Fantoni, G., Teloni, D., (2017). *Workers 4.0: skills, profiles and jobs in different business functions.* Economy, Employment and Skills: European, Regional and Global Perspectives in an Age of Uncertainty, 95-103, ISBN 9788895380452

Fernández-Sanz, L., Gómez-Pérez, J., Castillo-Martínez, A., (2017). *E-Skills Match: A framework for mapping and integrating the main skills, knowledge and competence standards and models for ICT occupations.* Computer Standards Interfaces,51, 30-42. doi:10.1016/j.csi.2016.11.004

Fondazione Giacomo Brodolini, Economy, (2017). *Employment and Skills: European, Regional and Global Perspectives in an Age of Uncertainty,* ISBN 9788895380452

Freddi D., (2017). *Digitalisation and employment in manufacturing.* Ai Society, 33(3), 393-403. doi:10.1007/s00146-017-0740-5

Frey C.B., Osborne M.A., (2017). *The future of employment: How susceptible are jobs to computerization?* Technological Forecasting and Social Change, Sept 1–72.

Fulmer, I. S., Ployhart, R. E. (2013). *Our most important asset.* Journal of Management, 40(1), 161–192. doi:10.1177/0149206313511271

Glas, A. H., Kleemann, F. C. (2016). *The impact of industry 4.0 on procurement and supply management: A conceptual and qualitative analysis.* International Journal of Business and Management Invention, 5(6), 55-66.

Gorecky, D., Schmitt, M., Loskyll, M., Zühlke, D., (2014). *Human-machine-interaction in the Industry 4.0 era.* In: In Proceedings of the 2014 12th IEEE International Conference on Industrial Informatics (INDIN), Porto Alegre, Brazil.

Grugulis, I., Vincent, S., (2009). *Whose skill is it anyway? 'Soft' skills and polarization.* Work Employ. Soc. 23 (4), 597–615.

Hamiz, M., Bakri, M., Kamaruddin, N., Mohamed, A. (2018). *Assessment analytic theoretical framework based on learners' continuous learning improvement.* Indonesian Journal of Electrical Engineering and Computer Science,11(2), 682.

Hankamer J., (1989). *Morphological parsing and the lexicon.* In Lexical representation and process, (pp. 392-408). MIT Press.

Hartmann E., Bovenschulte M., (2013). *Skills needs analysis for "Industry 4.0" based on roadmaps for Smart Systems,* in "Using Technology Foresight for Identifying Future Skills Need". Global Workshop Proceedings, edited by Moscow School of Management Skolkovo-ILO.

Haukka, S., (2011). *Education-to-work transitions of aspiring creatives.* Cultural Trends 20 (1), 41–64.

Hong L., Davison B.D., (2010). *Empirical study of topic modeling in Twitter*, Dept. of Computer Science and Engineering.

Hu M., Liu B., (2004). *Mining and summarizing customer reviews.* KDD '04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining Pages 168-177

ILO OECD, (11-12 June, 2018). *Approaches to anticipating skills for the future of work*, Section 3.4. Geneva, Switzerland.

Jerman, Andrej Pejic Bach, Mirjana Bertoncelj, Andrej. (2018). *A Bibliometric and topic analysis on future competences at smart factories.* Machines. 6. 41. 10.3390/machines6030041.

Kagermann H., Wahlster W., Helbig J., (2013). *Recommendations for implementing the strategic initiative INDUSTRIE 4.0.*

Kamprath, M., Mietzner, D., (2015). *The impact of sectoral changes on individual competences: A reflective scenario-based approach in the creative industries.* Technological Forecasting and Social Change.

Karakatsanis, I., Alkhader, W., Maccrory, F., Alibasic, A., Omar, M. A., Aung, Z., Woon, W. L., (2017). *Data mining approach to monitoring the requirements of the job market: A case study.* Information Systems,65, 1-6. doi:10.1016/j.is.2016.10.009

Khobreh, M., Ansari, F., Fathi, M., Vas, R., Mol, S. T., Berkers, H. A., Varga, K., (2016). *An ontology-based pproach for the semantic representation of job knowledge.* IEEE Transactions on Emerging Topics in Computing,4(3), 462-473.

Lado, A. A., Wilson, M. C., (1994). *Human resource systems and sustained competitive advantage: A competency- based perspective.* Academy of Management Review, 19: 699-727.

Lafer, G., (2004). *Conceptual confusions: emotion work as skilled work.* In: Warhurst, I.F.C. (Ed.), The Skills that Matter: Clinical Perspectives on Work and Organisations. Palgrave Macmillan, Basingstoke, pp. 109–127.

Lasi H., Fettke P., Kemper H.G., Feld T., Hoffmann M., (2014). *Industry 4.0.* Business Information Systems Engineering, 6(4).

Last C., (2017). *Global commons in the global brain.* Technological Forecasting and Social Change, 114, 48-64.

Lee, S., Yoon, B., Lee, C., Park, J. (2009). *Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping.* Technological Forecasting and Social Change,76(6), 769-786.
doi:10.1016/j.techfore.2009.01.003

Lloyd C., Payne, J., (2009). *Full of sound and fury, signifying nothing': interrogating new skill concepts in service work — the view from Two UK call centres.* Work Employ. Soc. 23 (4), 617–634

Lorentz, H., Töyli, J., Solakivi, T., Ojala, L., (2013). *Priorities and determinants for supply chain management skills development in manufacturing firms.* Supply Chain Manag. Int. J. 18 (4), 358–375.

Lumauag, R.G., (2019). *Decision support system for personnel selection.* International Journal of Recent Technology and Engineering 8 (1), 2019. 2019.

Ma, T., Zhang, Y., Huang, L., Shang, L., Wang, K., Yu, H., Zhu, D. (2017). *Text mining to gain technical intelligence for acquired target selection: A case study for Chinas computer numerical control machine tools industry.* Technological Forecasting and Social Change, 116, 162-180. doi:10.1016/j.techfore.2016.10.061

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., Barton, D. (2012). *Big data: the management revolution.* Harvard business review, 90(10), 60-68.

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.

Manad, O., Bentounsi, M., Darmon, P. (2018). *Enhancing talent search by integrating and querying big HR Data.* 2018 IEEE International Conference on Big

Data (Big Data). doi:10.1109/bigdata.2018.8622275

Mirski, P., Bernsteiner, R., Radi, D., (2017). *Analytics in human resource management: The OpenSKIMR approach.* Procedia Computer Science,122, 727-734. doi:10.1016/j.procs.2017.11.430

Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., Ueda, K., (2016). *Cyber-physical systems in manufacturing.* CIRP Ann. Manuf. Technol. 65 (2), 621–641.

Nadeau D., Sekine S. (2007), *A survey of named entity recognition and classification.* Lingvisticae Investigationes. 30. 10.1075/li.30.1.03nad.

Nam T., (2018). *Technology usage, expected job sustainability, and perceived job insecurity.*Technological Forecasting and Social Change. 138. 10.1016/j.techfore.2018.08.017.

Nesterenko, A., (2016). *Extracting functional job roles from professional social networking site profiles.* AIST.

Paris, I., (2012). *The Italian white-goods industry and the European Common Market during the years of the 'economic miracle' (1958–63): quantitative evidence and interpretative hypotheses.* European Review of History: Revue européenne d'histoire, 19:4, 575-599, DOI: 10.1080/13507486.2012.697873

Pejic Bach, Mirjana Bertoncel, Tine Meško, Maja Krstić, Živko. (2019). *Text mining of industry 4.0 job advertisements.* International Journal of Information Management. 10.1016/j.ijinfomgt.2019.07.014.

Piatetsky-Shapiro, G., (1991). *Discovery, analysis, and presentation of strong knowledge discovery in databases.* AAAI/MIT Press, Cambridge, 248, 255-264.

Poonnawat, W., Pacharawongsakda, E., Henchareonlert, N. (2018). *Jobs analysis for business intelligence skills requirements in the ASEAN Region: A text mining study.* Advances in Intelligent Systems and Computing Advances in Intelligent Informatics, Smart Technology and Natural Language Processing,187-195. doi:10.1007/978-3-319-94703-7_17

Porter, M.E., Heppelmann, J.E., (2014). *How smart, connected products are transforming competition.* Harv. Bus. Rev. 92 (11), 64–88.

Pryima, S., Rogushina, J.V., Strokan, V., (2018). *Use of semantic technologies in the process of recognizing the outcomes of non-formal and informal learning,* (2018) CEUR Workshop Proceedings, 2139, pp. 226-235.

R Core Team, (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Right, G. S. (2016). *Assessing and anticipating changing skill needs.* OECD Publishing.

Rosenberg M., (2009). *The surprising benefits of robots in the DC.* Supply Demand Chain Executive, 10 (2): 39-40

Rotman D., (2013). *How technology is destroying jobs.* MIT Technology Review, June 12. http://www.technologyreview.com/featuredstory/515926/how-technology-is-destroying-jobs/

Roy, R., Stark, R., Tracht, K., Takata, S., Mori, M., (2016). *Continuous maintenance and the future – Foundations and technological challenges.* CIRP Annals - Manufacturing Technology, 65 (2), pp. 667-688. DOI: 10.1016/j.cirp.2016.06.006

Sawleshwarkar, S., Rangnani, N., Mariwalla, V., Halbe, A., (2018). *Simplified recruitment model using text-mining on psychometric and aptitude tests.* 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). doi:10.1109/iceca.2018.8474769

Selvam M., Gayathri J., Vasanth V., Lingaraja K., Marxiaoli S., (2016). *Determinants of firm performance: A subjective model.* International Journal of Social Science Studies, 4(7). doi: 10.11114/ijsss.v4i7.1662

Shah, R., Ward, P. T., (2003). *Lean manufacturing: context, practice bundles, and performance.* Journal of operations management, 21(2), 129-149.

Sung, T. K., (2018). *Industry 4.0: A Korea perspective.* Technological Forecasting and Social Change,132, 40-45. doi:10.1016/j.techfore.2017.11.005

Szalavetz, A., (2018). *Industry 4.0 and capability development in manufacturing subsidiaries,* Technological Forecasting Social Change,
https://doi.org/10.1016/j.techfore.2018.06.027

Tambe, P., Ye, X., (2018). *Sourcing and screening high-tech workers.* ICIS (2017): Transforming Society with Digital Innovation

Tarry, A., (2019). *Coaching with careers and AI in mind: Grounding a hopeful and resourceful self fit for a digital world.* Abingdon, Oxon, England: Routledge.

Thoben K.-D., Wiesner S.A., Wuest T., (2017). *Industrie 4.0 and smart manufacturing - a review of research issues and application examples.* International Journal of Automation Technology, 11(1)

Thompson, K. R., Koys, D. J., (2010). *The Management Curriculum and assessment Journey: use of Baldrige criteria and the occupational Network database.* Journal of Leadership Organizational Studies, 17(2), 156–166. doi:10.1177/1548051810369341

Tseng Y.-H., Lin C.-J., Lin Y.-I, (2007). *Text mining techniques for patent analysis.* Information Processing and Management.

Tunzelmann, N. V., (2003). *Historical coevolution of governance and technology in the industrial revolutions.* Structural Change and Economic Dynamics,14(4), 365-384. doi:10.1016/s0954-349x(03)00029-8

Weber, E., (2016). *Industry 4.0: Job-producer or Employment-destroyer?* Retrieved on March 2017, http://doku.iab.de/aktuell/2016/aktueller_bericht_1602.pdf.

Wickham H., Hester J., Francois R., (2018). *readr: Read Rectangular Text Data.* R package version 1.2.1. https://CRAN.R-project.org/package=readr

Wilson R., (2013). *Skills anticipation - The future of work and education.*

World Economic Forum, 2016. The Future of Jobs. http://www3.weforum.org/docs/WEF_FOJ_Executive_Summary_Jobs.pdf

Zhao, M., Javed, F., Jacob, F., McNair, M., (2015). *SKILL: A system for skill identification and normalization.* AAAI..

3 Skill(N)er: Mining and Mapping Soft Skills from Any Text[14]

3.1 Introduction

Academia and industry have shown a growing interest on soft skills in the last few years. The focus of scholars and practitioners on this topic has grown for many reasons, but the main one is (counterintuitively) digitalisation. It is known that the impact of digitalisation is widespread (Van Laar et al, 2017) and heterogeneous (Galati et al., 2017), but is also evident that many studies outline the importance of acquiring soft skills in order to face the digital wave (Chryssolouris et al., 2013; Gorecky et al., 2014; Weber, 2016), stressing that a competence gap in this context will have a negative impact on the workforce (Frank et al., 2019; Bauer et al., 2011; Bridgstock, 2011; Dobrunz et al., 2006; Haukka, 2011; Cooper and Tang, 2010). The fear of robotisation seems to be the main driver of the increasing focus on soft skills. As (Fernbach Sloman, 2017) stated, machines cannot share their intentions, they have not consciousness, they are not able to influence others and they do not collaborate with humans; in other words, humans merely use technology. Moreover, (Hanskel & Westlake, 2018) outlined that power in the workforce of the future would be in the hands of professional profiles characterized by a combination of noncognitive skills (especially personal interaction) and cognitive skills, able to manage intangible assets. Since knowledge assets are frequently contestable, companies are more and more searching for people able to do it and, most of all, to identify and maximize synergies.
Another emblematic work (Frey et al., 2017) estimated that around 47% of jobs are at high-risk of robotisation, especially the ones characterized by routine tasks. Anyway, the results look encouraging for job profiles characterized by a high soft-content, since the professional profiles assigned to the low-risk category are the ones characterized by an higher presence of soft skills. Indeed many scholars worked in the last years on soft skills looking for innovative ways to facilitate their development (Sanz et al., 2019; Tseng et al., 2019; Duran-Novoa et al., 2011), their assessment (Bohlouli, et al., 2017), their comprehension (Chechurin et al., 2016) or more specifically identifying how they impact digital jobs (Hendon et al., 2017). Despite this increasing focus, what emerges is that there are many relevant limitations to the current approaches and gaps in the literature regarding the study of soft skills:

- approaches are mainly qualitative thus human labour intensive and hardly replicable;

[14] *Fareri, S., Melluso, N., Chiarello, F., Fantoni, G., (2020). "Skill(N)ER: Mining and Mapping Soft Skills from any Text", Expert Systems with Applications (under review)*

- it is hard to converge (with a top-down approach) to a common definition of soft skills (Robles, 2012; Mitchell, 2010; Schultz, 2008; Evenson, 1999);
- it is hard to create (with a bottom-up approach) a complete list of all the skills that can be considered "soft";
- it is not clear if (and eventually what kind) a relationship exists between soft skills (e.g. in the way soft skills are shared by job profiles) (Hendon et al., 2017).

These limitations make it hard to increase our knowledge in the context of soft skills. In the present paper, we propose a method based on Natural Language Processing (NLP) in order to overcome these gaps. The method has the goal of automatically collecting soft-skills from documents and investigating the existing relationships between these skills. We relied on a specific class of NLP algorithms, Named Entity Recognition (NER) or Entities Retrieval Approaches (Nadeau et al, 2017; Etzioni et al, 2005; Ritter et al, 2011) systems. These algorithms are able to collect all the entities of a pre-defined type contained in the documents. In order to understand the value of the present paper, the reader has to take into consideration that it is not a straightforward task to apply standard NER systems for the automatic extraction of soft-skills, since state of the art NER systems are based on supervised classifiers. For a supervised classifier to work, many (i.e. tens of thousand) examples of manually tagged entities are needed: this is why these algorithms perform well on tagging entities such as names of people, places and products. In the case of soft skills (a rare and ill-defined entity), creating high-quality manually annotated corpus requires the work of experts in the field (with the many limitations listed above). In order to build a reliable entity recognition system it is crucial to define a validated list of soft skills and to take into account the linguistic context in which each skill appears. For this reason, we use the following approach: we exploit experts' consultation to build linguistic rules to define a list of clues (words indicating the presence of a soft skill) and a list of soft skills. These two lists are the constituent elements of what we have called Extraction Context, sentences and fragments of sentences containing (with a high likelihood) a soft skill. The Extraction Context is what we use to build our NER system, the SkillNER. This approach helps to increase the precision and recall of our entity extraction system and to assess the quality of the output. We further confronted our system with a system built over the ESCO[15] Transversal Skills. This analysis demonstrated that for the extraction of soft skills from any text it is crucial to understand the linguistic context in which a skill appears. The paper is structured as follows: firstly, we give an overview of the existing studies on soft-skills and of the Competence Classification Systems together with analyses which have already been performed on such databases. Secondly, the

[15](European Skill/Competence Qualification and Occupation)

methodology adopted to reach the objective of building an entity extraction system for soft skills. Then, we show a graph based analysis of soft skills carried out on the ESCO (European Skills/Competence Qualification and Occupation) database. At last, we discuss the limitations of the proposed approach and possible future developments focusing on potential stakeholders and fields of application of the proposed system.

3.2 Literature Review

Soft skills are considered ever more relevant in a world focused on digitalization; unfortunately, the topic lacks scientific methodologies to be studied. Through the prior art, the authors try to explain why, giving also an overview of the databases and algorithms subsequently adopted. The state-of-art is structured as follows: first of all, the authors give demonstrations of the lack of a common definition of soft skills (which is definitely a signal of the complexity of the domain analyzed), and the markedly increased interest on the topic, visualizing the trend of citation on the last 20 years. Then, they give an overview of the main skill databases and the analyses performed considering them as primary sources. Finally, a brief description of Named Entity Recognition (NER) algorithms is presented.

3.2.1 Soft skills: a hot Topic, a complex Domain

Soft skills not only have a key role for withstanding the penalizing effects of computerization on employment (Frey, 2017), but they are ever more considered a critical factor of success (Hungwei Tseng et al., 2018; Deming D. et al., 2017) . Although some researchers tried to make order, conceptualizing soft skills as constructs made by other sub-skills (Blake et al., 2011), they are still a hard theme to study since there are no defined boundaries that delineate the domain. Since there is no common naming for soft skills, different labels have been proposed by Institutions and Academia (a signal that the topic is still hazy). In table 8, we show some of the most reliable ones. (Cinque, 2015)

Furthermore, the complexity is even more accentuated due to the fact that there is no universally recognized definition in literature, thus making it hard to understand its meaning. As evidence of such fuzziness, in table 9 we report the most cited references on Scopus (November 2019) regarding the topic and the definitions they provide.

Our best knowledge shows partially overlapping (and in some cases, contradictory) definitions of soft skills. It is interesting to notice that also the same author gives different formal definitions of the same concept, even if with a similar meaning. If the previous phenomenon occurs in a single author, it is easy to imagine how heterogeneous other definitions can be. Some authors outline the embedded

Soft Skills Naming	Source
Life skills	(WHO, 1993)
Transversal skills	(ISFOL, 1998)
Generic Competences	(Tuning Project, 2000)
Key competences for a Successful life and a well functioning society	(OECD, 2003;2012)
Key competences for lifelong learning	(UE, 2006)
21st centuries skills	(Ananiadou & Claro, 2009)
Transferable skills	(RPIC-VIP, 2011)
Future Work skills	(IFTF, 2010)
Soft skills for Talent	(Manpower Group, 2014)
Skills for social progress	(OECD, 2015)

Table 8: Heterogeneous naming of soft skills from the most reliable references (Cinque, 2015)

Soft Skill Definition	Source	Citations
Soft skills are interpersonal qualities, also known as people skills, and personal attributes that one possesses.	Robles, 2012	907
Soft Skills = Interpersonal (People) Skills + Personal (Career) Attributes.	Robles, 2012	907
People skills promote a positive attitude, effective communication, respectful interaction, and the ability to remain composed in difficult situations	Evenson, 1999	407
Soft skills refer to the cluster of personality traits, social graces, facility with language, personal habits , friendliness, and optimism that mark people to varying degrees.	Schulz, 2008	353
Soft skills describe career attributes that individuals should possess, such as team skills, communication skills, ethics, time-management skills, and an appreciation for diversity.	Mitchell, 2010	89

Table 9: Heterogeneous definitions of soft skills from the most cited references

aspects of soft skills such as character traits and certain forms of employment that do not depend on acquired knowledge; other researchers give major light to interpersonal abilities and the capacities of having efficient relationships with colleagues. Although the theme is definitely fuzzy, we lean towards (Robles, 2012) definition, since it seems to be the widest one, capturing the complex nature of soft skills that we agree to be hybrid and made by two macro-classes: the soft skills strictly related to the single individual and the soft skills necessary to optimize relationships with others. Nevertheless, even if it is difficult to define and measure, the extreme contemporary relevance of soft skills is undoubtedly evident. To demonstrate to what extent the interest in the topic is growing, we search for the keyword "soft skill*" on all the fields of Scopus papers (Date: 28/12/2019). We intentionally considered all fields with no particular research restrictions in order to have the whole picture of such a topic. The query returns 7539 papers. The distribution has a well-defined growing trend; in particular, academic papers show a peak of interest after 2014, the same year of release of The Second Machine Age (Brynjolfsson et al., 2014) in which the authors argued over the automation of cognitive tasks and explaining that humans and machines are perfect substitutes for one another and do not complement each other.

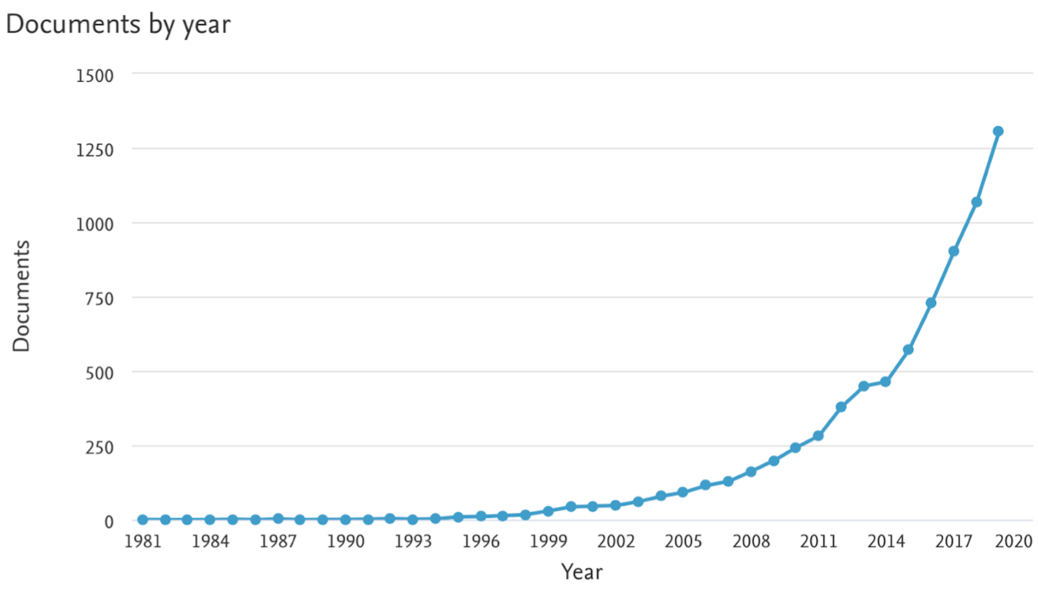


Figure 11: Trend of publication on "soft skills" (all fields) from 1981 to 2019.Source: Scopus. Date: 28/12/2019

Also the distribution of documents that deals with soft skills (represented per research field) is extremely meaningful. We would expect a predominance of papers linked to social science and psychology, instead there is also a high percentage which is related to engineering and computer science that deserves attention. In Fig. 12 we show the distribution of papers per research field from 2000 to 2019. As we can see, social science is the most important area, even if its relevance seems to decrease after 2010 and it is now making the way to an even bigger number of researches belonging to business. Furthermore, the theme is increasingly treated in Arts and Humanities but also in Mathematics, which was totally missing at the beginning of the considered period. Finally, the heterogeneity of fields, proves the interdisciplinarity of the topic and the subtended heterogeneous stakeholders involved.

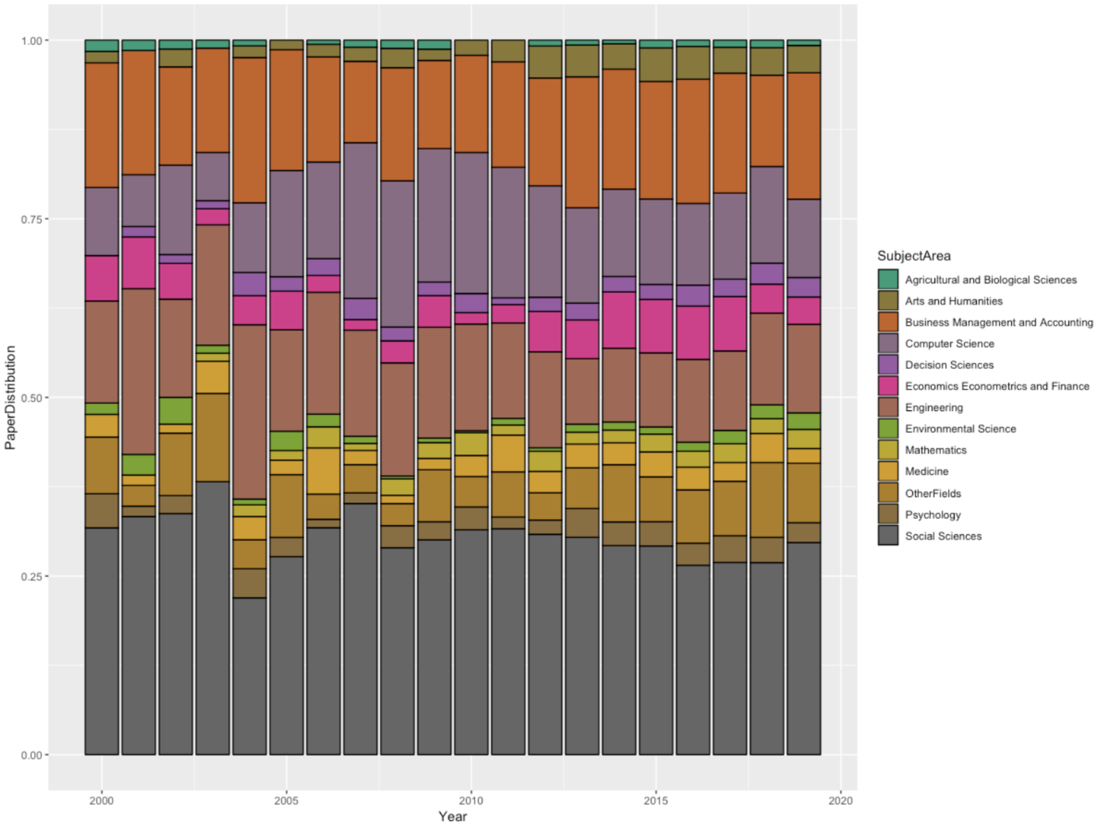


Figure 12: Breakdown of Soft Skills papers from 2000 to 2019 per research fields.

3.2.2 The Taxonomies of Skills

Another approach to define soft skills from the bottom, is listing them. A remarkable effort in this direction has been done by Taxonomies of Skills. Several taxonomies of skills can be found online, but the primary sources of occupational information are ESCO[16] (European) and O*NET[17] (American). ESCO (European Skill/Competence Qualification and Occupation) is a multilingual classification system for Europe; it classifies jobs, capabilities, competences, and qualifications in Europe that are relevant for the labor market. Through a triangular relationship among skills, profiles, and qualifications, the aim of ESCO is bridging the gap between academia and industry in all of Europe.

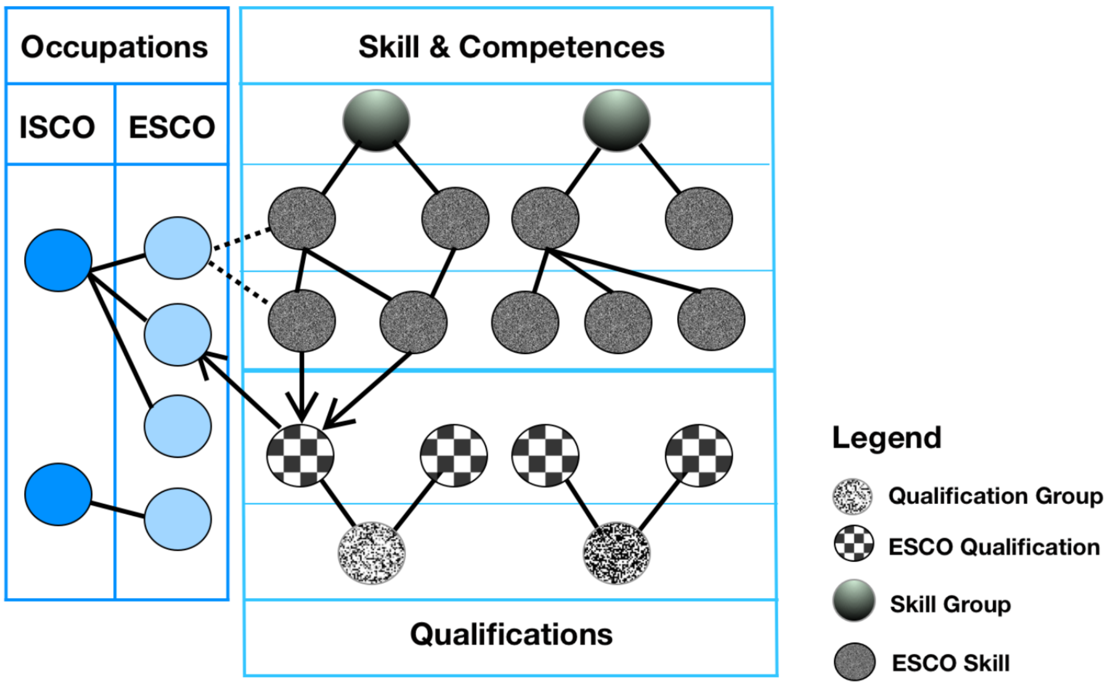


Figure 13: The Three Pillars approach. The relationship between groups of skills/qualifications and single elements is hierarchical. One ISCO occupation could correspond to multiple ESCO occupations or to a single one. Each ESCO occupation is characterized by a heterogeneous number of skills (essential or optional). The database shows which is the necessary learning path to acquire a specific qualification and which qualifications are relevant for each ESCO occupation.

The occupation classification corresponds to ISCO-O8, which is the International Standard Classification of Occupations (International Labor Organization, 2008). ESCO is made up of 2942 occupations, 13485 skills, 9606 qualifications and, most of all, their relations. Occupations, skills, and qualifications are all organized into a hierarchy. ESCO structure is based on three pillars (occupations, skills, and

[16]https://ec.europa.eu/esco/portal/home
[17]https://www.onetcenter.org

qualifications) that are interlinked to show the relationships between them; it also includes information about whether skills and competencies are essential or optional and what qualifications are relevant for each ESCO Occupation. Moreover, it has been developed in an open IT format and it is open source and easy to be collected. O*NET is the American correspondent of ESCO. O*NET is an available online database developed for the U.S Department of Labor which is made of 974 occupations from Standard Occupational Classification (SOC) and their corresponding skills, knowledge, and abilities. The use of SOC makes it possible to analyze professions from multiple perspectives, comparing data from different federal sources, aggregating data on employment from a wide range of job titles, and tracking data over time to identify changes in the labor market. The taxonomy contains abilities, occupational interests, work values, work styles, basic skills, cross-functional skills, domains of knowledge and items related to prior educational experience required to perform in a job, skills related to experience requirements, skills related to occupational requirements and skills related to occupation-specific information. It also contains the tasks and the "tools and technologies" for each occupation. Each job profile has quantitative information about level and importance for every owned skill described above. The conceptual foundation of the O*NET framework is represented in figure 14, which shows the most important information contained in the database.

Feature	ESCO	O*NET
N. of Skills or Descriptors	13485	277
N. of Occupations	2942	974
Qualifications	9606	0
Job Classification	ISCO08	SOC
Format	.xls, .txt, MySQL, .RDF	.RDF, .csv, .xls

Table 10: The main quantitative (and qualitative) differences between ESCO and O*NET.

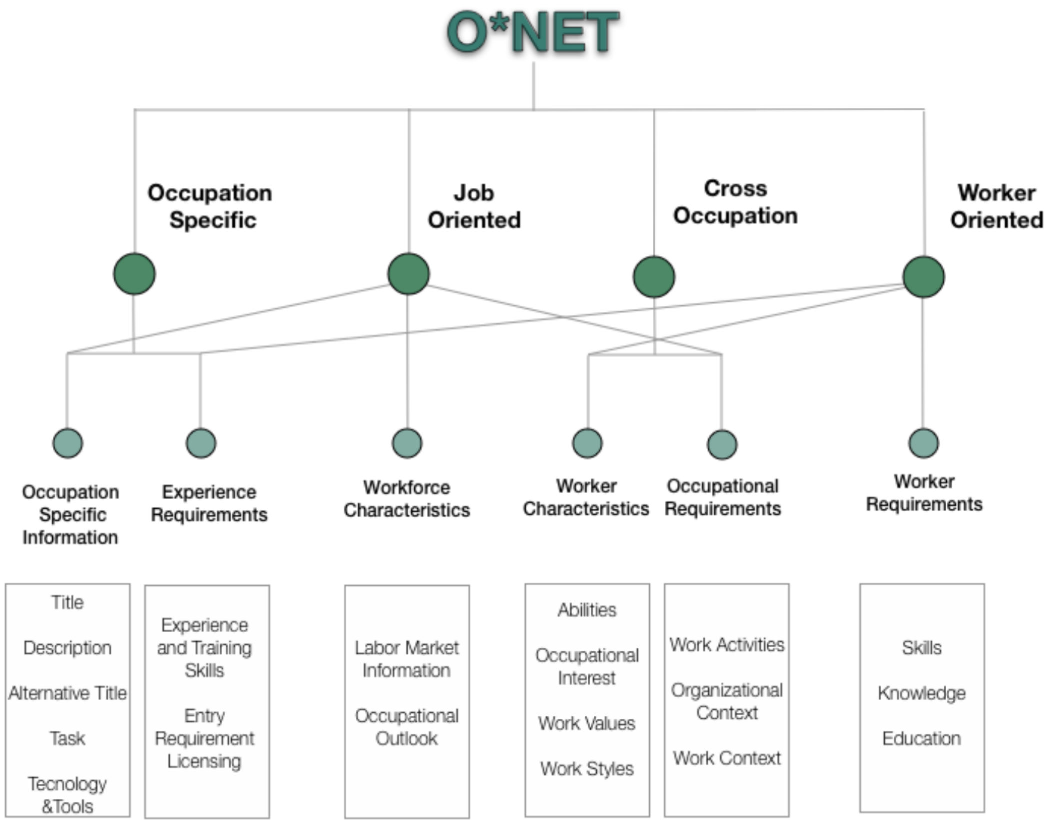


Figure 14: The O*NET content model. https://www.onetcenter.org/content.html)

To summarise, the main quantitative and qualitative differences between the two taxonomies are shown in table 10.

As we can deduce from both Fig. 13 and Fig. 14, the granularity of the occupations (and skills) is extremely different, as can be inferred from the values in table 10. In contrast to O*NET, ESCO contains almost 10000 qualifications defined as the assessment and validation of individual learning outcomes; they are owned and managed by the European Member States. ESCO also has a greater level of detail than O*NET, having 6 times the number of skills and 3 times the number of job profiles (of O*NET). Furthermore, ESCO contains a large number of heterogeneous skills, which are frequently assigned only to a single job profile. On the contrary, O*NET has all the descriptors of figure 14 assigned to every job profiles, which differ only for their importance and level values (ranging from 0 to 100). Finally, there is not clear distinction between hard and soft skills in O*NET, while around 110 skills are labelled as transversal[18] in ESCO (v1.0.3); however, also ESCO has some weakness, because its soft skills are too abstract to be assigned to a specific job profile[19] and their number is still limited.

3.2.3 The Use of ESCO and O*NET as data-sources

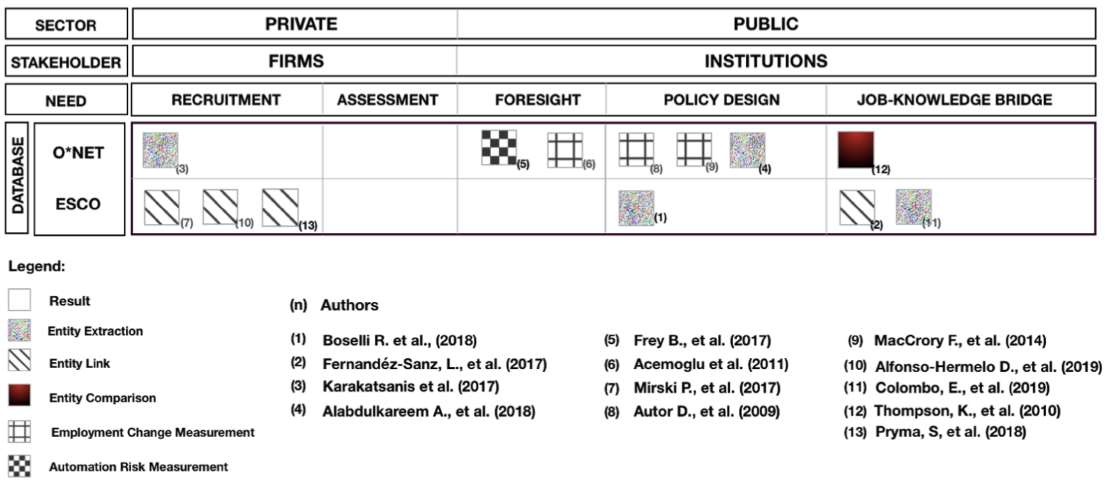


Figure 15: Literature review map, representation by authors. The map should be read as follows: the author ”n” responds to the need of the stakeholder ”k”, who belongs to sector “m”, offering ”p” as the main result of his analysis and as a possible support/response to his need, analyzing the database “o”.

ESCO and O*NET, as shown in figure 15, are frequently used by researchers to extract information on skills and job profiles. The main insights shown in the proposed map are:

- a group of papers reports results about the detection of employment drivers of change and the measures used to define the susceptibility of workers to

[18]File transversalSkillCollection.csv
[19]https://ec.europa.eu/esco/portal/document/it/87a9f66a-1830-4c93-94f0-5daa5e00507e

automation. The techniques addressed in these papers belong to the field of econometrics (Frey et al., 2017; Arntz et al., 2017; Autor et al., 2009; Acemoglu et al., 2011; MacCrory et al., 2014);

- other papers focus on the development of further skills taxonomies or entity extraction (Boselli et al., 2018; Karakatsanis et al., 2017; Alabdulkareem et al., 2018; Colombo et al., 2019), data comparison (Thompson et al., 2010) and data linking (Mirski et al., 2017; Fernandéz-Sanz et al., 2017; Alfonso-Hermelo et al., 2019, Pryima et al, 2018);
- the approaches for predictive purposes (Frey et al., 2017; Acemoglu et al., 2011) and effective policy design (Alabdulkareem et al., 2018; Autor et al., 2009; MacCrory et al., 2014) are mainly founded on O*NET, maybe because its level of detail (ESCO has a finer grain of occupations and skills) suits better the level of detail required for econometrics. The difference between the econometric analysis and recruitment related problems study is that the first one studies a global behavior and it needs a database that has a low level of detail; on the other hand, the second case requests a micro analysis;
- there exists a frequent (Mirski et al., 2017; Alfonso-Hermelo et al., 2019; Pryima et al, 2018) attempt of improving the effectiveness of the recruitment process thanks to the linkage of skills to the ESCO framework;
- there are several attempts to build bridges between universities and market requirements, particularly through text mining techniques (Fernandéz-Sanz et al., 2017; Colombo et al., 2019; Thompson et al., 2010);
- to our best knowledge, no research about firms' skill assessment developed through ESCO and O*NET exists; this may be due to the fact that the process is made internally and the value obtained is not shared.

Overall, the table contains works that respond to the needs of the public sector (with respect to the private one). This can be due to two main facts:

1. work and jobs related information changes from company to company;
2. there exists a reluctance of private sector to publish this kind of data.

What seems to be missing (and critical for the authors) is a solution which responds to multiple stakeholders and that is source independent, so that it can be tested also on both on public and private data.

3.2.4 Named Entity Recognition

As stated in section 3.1, in the present paper we try to solve the lack of methodological and data-driven approaches for soft-skills identification using Natural Language Processing (NLP), and in particular Named Entity Recognition (NER) Tools (Harrag, 2014). Information extraction from unstructured documents is a task that has been subject of active research for several years in the community of Artificial Intelligence (Harrag, 2014) and Computer Engineering (Choudhary et al., 2009). Current approaches focus on processing text in order to make a meaningful representation of concepts (and their relationships) contained in documents (Piskorski, 2013; Gildea, 2014; Liu et al., 2020). In particular, NER consists of detecting lexical units in a word sequence that refers to a predefined entity, thus determining what kind of entity the unit is referring to (e.g. persons, locations, organizations.). The methods used for NER are various:

1. terminological-driven NER: aims to map mentions of entities within texts to terminological resources (e.g. wikipedia) (Nadeau et al, 2017);

2. rule-based NER: uses lexicons, regular expressions and lexical information to express knowledge based systems able to extract a certain type of entity (Sari et al, 2010);

3. corpus-based NER: uses manually tagged text corpora (training set) to train machine learning (ML) algorithms (Quinlan, 1986; Suykens, 1999; Lafferty et al, 2001; McCallum et al, 2010)

NER has been successful in different languages and different domains. It can provide crucial, although shallow, semantic information for tasks such as question answering (Abujabal et al, 2018; Blanco-Fernández et al., 2020), topic disambiguation (Fernández, N. et al. 2012) or detection (Krasnashchok et al, 2018; Lo et al., 2017, Al-Nabki et al., 2019) and revealment of elements relationships (Sarica et al., 2020; Amal et al., 2019). Furthermore, NER has proved to be effective in broader applications, such as user profiling (Nicoletti et al., 2013) and ontology development on unconventional domains (Oliva et al., 2019; Rodrigues et al., 2019). Anyway, since NER is a classification task, for using the most advanced approach in terms of accuracy (corpus-based NER uses deep neural networks) (Devlin et al, 2018), a training set is needed, i.e. a set of manually annotated documents. Since manually annotated data is rarely available, there are only a few types of entities that can be mined with high accuracy, even by the most sophisticated deep learning systems. Applying NER systems trained to extract a specific entity in a specific domain (e.g. names of persons from scientific papers) to extract another kind of entity from a novel domain (e.g. names of cities from Twitter) yield to a certain failure (Ciaramita et al, 2005). This problem is increased due to the partial inconsistencies in the manual tagging procedure, especially in domains where manual tagging requires very specific expertise (i.e. the soft-skill one). Thus at the state of the art, NER systems could be considered effective only if the system is designed by experts in NLP and domain experts (Chiarello et al, 2018).

3.3 Materials and Methods

The goal of this study is to develop a system that automatically detects and extracts soft skills from text. We deal with this objective by building a Named Entity Recognition system. In Section 3.3.1 we describe the linguistic features that the Named Entity Recognition system uses; in Section 3.3.2 we focus on describing how soft skills appears in the text introducing the Extraction Context; in Section 3.3.3 we report the methodological steps undertaken the design of our system, in particular, focusing on the Clue Extraction, Skill Extraction and Training.

3.3.1 Linguistic Features for Named Entity Recognition

A rule-based Named Entity Recognition system relies on using a set of rules defined by human experts to extract entities. For example, one rule could be expressed as "if a proper noun follows a person's title, such as 'Mr', then the proper noun is a person's name". Hence this approach takes a set of pattern that consists of the combination of two elements:

- lexical, grammatical and syntactic features;
- defined lists of words.

In the case of the detection of the name of a person the rule takes into account the grammatical feature of a word (proper noun) combined with the word "Mr". Also a corpus-based NER system uses the linguistic features of a text to take advantage of the machine learning algorithms: these algorithms are trained using sets of features extracted from annotated datasets.

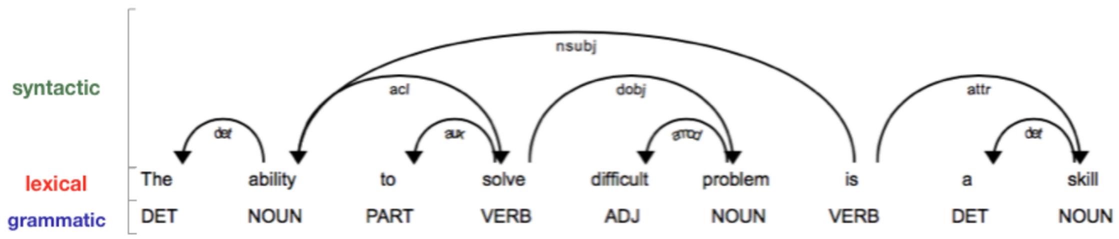


Figure 16: Example of linguistic features of a sentence

Figure 16 shows an example of linguistic features of a sentence. The lexical feature represents the sentence as a list of tokens (contiguous characters between two spaces, or between a space and punctuation marks). The grammatical feature represents the sentence by the part-of-speech of each token. Finally, the syntactic feature, which is obtained by using a dependency parser, represents the logical relationship between the words. The syntactic layer can be represented with a tree structure, where the words are connected to each other by arcs that are labelled by the name of the relation. This is particularly important for our process because soft skills can also be noun chunks. Noun chunks are base noun phrases, flat phrases that have a noun as their head. In short, it is possible to think of noun chunks as a noun plus the words describing the noun – for example, “lavish green grass” or “world’s largest tech fund”. We define the rules of our system relying on these linguistic annotations. The rules are defined as patterns to identify words, phrases and relationships according to the following syntax:

- LEMMA: access to the lexical feature of a token
- POS: access to the grammatical feature of a token
- DEP: access to the syntactic feature of a token

With this syntax we access either a single or a multiple linguistic feature. Here we have an example of pattern:
[”LEMMA”:”active”,”POS”:”VERB”]

In this case, we are searching for an entity made of two words, whose first one has the lexical feature equal to “active”, while the second one has the grammatical feature equal to a verb. Namely, with this rule we capture phrases such as “active listening”, “active learning” or “active practicing”. The power of extracting and combining these features from the text is given by the “Rule-based matching engine” of Spacy (Honnibal et al. 2017), that is an open-source software library for Natural Language Processing, written in Python and Cython. This tool gives access to the tokens within the document and their relationships using the pattern above mentioned. In section 3.3 we show in detail how we define our patterns used to build our Named Entity Recognition model.

3.3.2 Extraction Context of Soft Skills

Building a NER system that extracts soft skills requires the study of the behavior of these entities when they appear in the text. The first problem was understanding which are the structured and recurrent patterns to detect them. In this study, we started from the hypothesis that a soft skill frequently appears preceded by recurrent textual patterns. The combination of the latter and the soft skill embodies what we call extraction context. The extraction context consists of the following elements:

- The Entity: a linguistic sign (i.e., one word or a set of words), that makes reference to the concept;
- The Clue: a set of terms, lexical expressions or recurrent patterns correlated with the appearance of the entity .

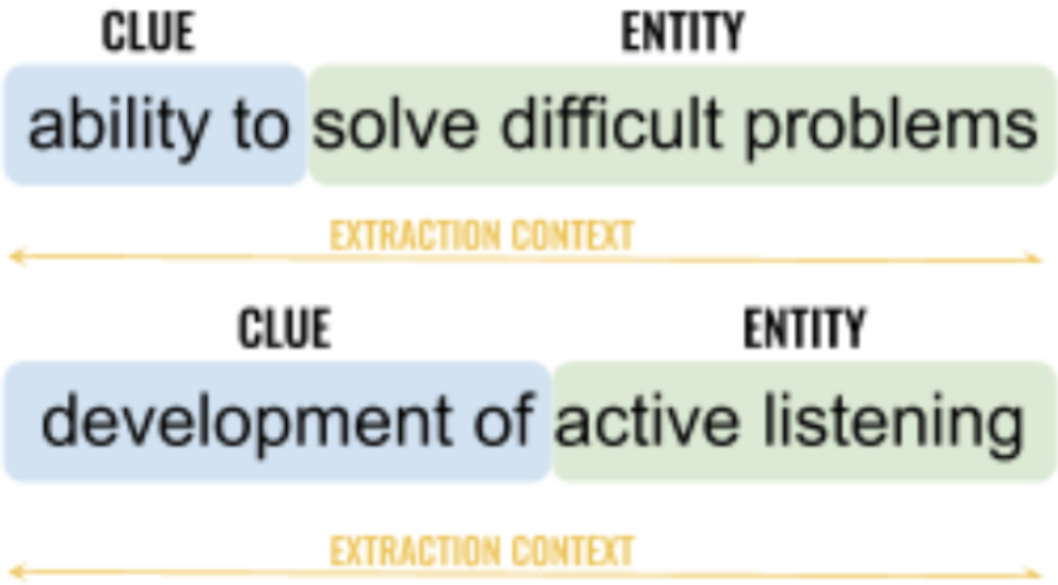


Figure 17: Examples of extraction contexts

Figure 17 shows an example of two extraction contexts that represent the clues and their associated entities.

The example in Figure 17 suggests that if in a sentence appears “ability to” or “development of” we have high probability that what is following is a soft skill. However, we know that a skill cannot always be preceded by a clue. Therefore, we can say that not all the extraction contexts contain a clue: very often the extraction context coincides with the skill itself. For example, in the sentence “the evaluation of participants is based on the assessment of their level of critical thinking and problem solving” we have two extraction contexts made of “level of critical thinking” and “problem solving”. In the first we have the clue “level of”, while in the second no clue appears. The definition of clue, namely a piece of information used to detect a certain hidden entity, is crucial for our work. If we know the recurrent patterns in the text which can usually refer to a soft skill we can define a system that is able to deal with the contextual information. The knowledge of the context in which the skills appear allows us to produce a system that is able to better discriminate the entities in the text. In the following section we will explain in detail how we use the extraction context to build our SkillNER.

3.3.3 Methodology

In this study we approach the Soft Skills Extraction problem by using a semi-supervised method. First, we use a rule-based approach to build a list of clues and a list of soft skills. Second, we use these two lists to train a supervised model based on a corpus in which the extraction context is annotated. The annotation of the corpus has been done automatically after the definition of the two lists. The methodology is divided into the following phases:

1. Clue Extraction: in this first phase we extract and validate a list of clues;
2. Skill Extraction: in this second phase, using as input the validated list of clues, we extract and validate a list of soft skills;
3. Training: in this phase we train a supervised model on a corpus annotated with clues and skills.

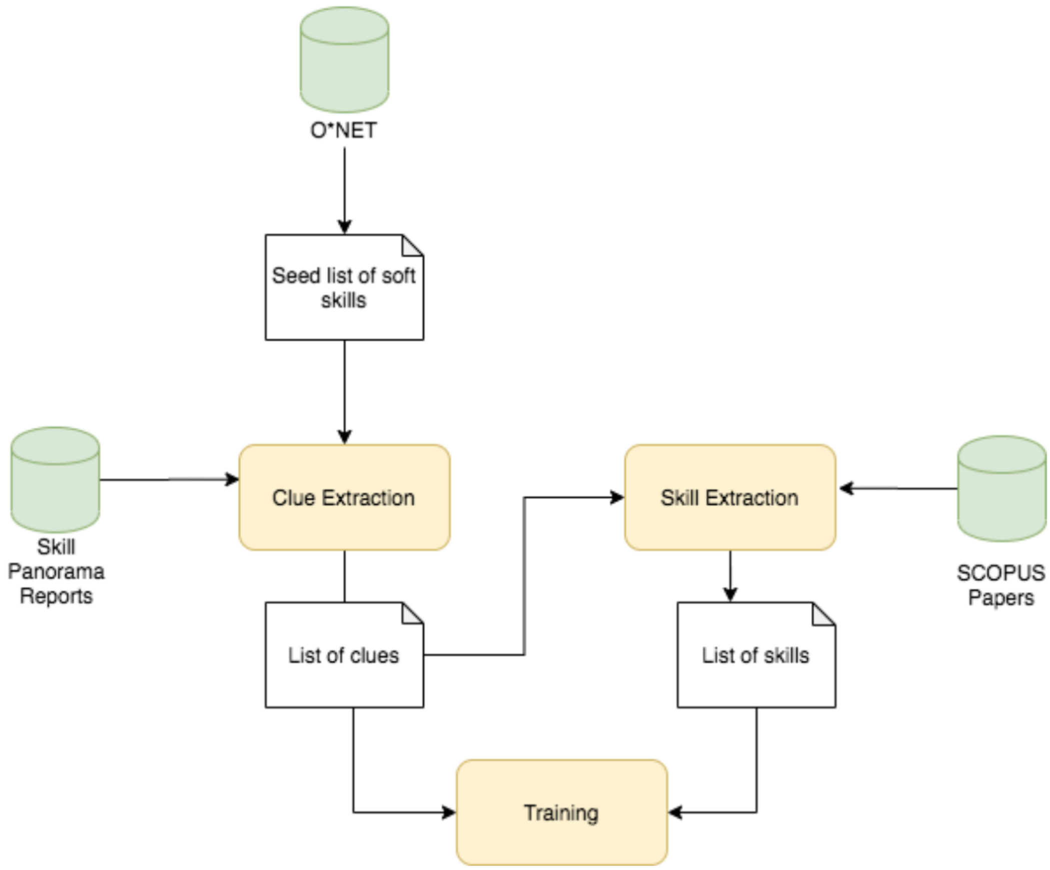


Figure 18: Methodological steps to build a Soft Skills Extraction System

Fig. 18 shows the flow diagram where four different elements are graphically displayed: activities (rectangular shape), check points (diamond shape), documents created from the procedure (sheet of paper shape) and databases (database shape).

Clue Extraction The aim of this phase is to extract a list of clues. First, we selected a small seed list of soft skills. Next, we look for these skills in a set of documents and see which words most frequently surround these skills. Finding a seed list is not a trivial task. We start performing a manual collection of soft skills, searching among the following sources:

- the three most cited papers on soft skills according to Scopus, whose reference on the topic was explicit in the title, abstract and keywords of the papers (the extraction was made on Scopus Database in November 2019) writing the following query: "TITLE-ABS-KEY ("soft* skill*" OR "social* skill*" OR "commun* skill*" OR "person* skill*" OR "languag* skill*")";

- the O*NET database, since it contains occupational definitions and skills, both hard and soft, from American workforce.

The concurrent analysis on O*NET has several advantages. First of all, the authors performed a cross-validation process: we identified the most cited soft skills in literature that are simultaneously present in an Occupational Framework. Moreover, since O * NET labels consist of a maximum of 3 terms, this step also guarantees to have concise formulations of skills, which are therefore more easily traceable in the text. We decided to search and extract clues among reliable documents belonging to the reports of "Skill Panorama". The Skills Panorama is an online central access point for data on skill needs in countries, occupations and sectors across EU Member States. It is an initiative of the European Commission aiming at improving the EU's capacity to assess and anticipate skill needs, helping education and training systems be more responsive to labour market needs, and better match skill supply and demand across the EU. We made use of a Rule-Based that allowed us to define rules that can be used to identify phrases from the documents. The system is based on the definition of the following simple rule:

["POS":"NOUN","OP":"*","LEMMA":"problem","LEMMA":"solving"]

For example, this rule extracts the first noun that precedes the skill "problem solving". In practice we are assuming that a clue is made of all the words that precedes a skill up to the noun. The output of the system is a list of phrases in which appears at least one of the soft skills of the seed list. The higher is the frequency of a clue, the higher is the probability that it precedes a soft skill. We choose our final list of clues using a twofold approach: first, measuring their frequency, second, by a critical assessment based on our expert judgment.

Skill Extraction The aim of this phase is to extract a list of soft skills. The investigation of the Extraction Context is also the key to accomplish this objective. Using the list of clues defined in the previous phase, we searched for them in a set of 5000 papers from Scopus and we collected all the sentences in which at least one clue appears. More specifically, we use the content of abstract of scientific papers where the word "soft skills" appears in the title, abstract or keyword. The results of academic studies are the most suitable place where to find skills. We choose to rely on scientific production because it is an authoritative source of reliable information with respect to other widely used sources (e.g. wikipedia, news or social networks). Each of these sentences was analyzed by a group of experts who looked for the soft skills contained in it. The output of this phase is a list of soft skills extracted manually from a set of phrases detected automatically. In fact, the knowledge of the list of clue helped the experts by reducing their efforts in terms of time; instead of analyzing the whole corpus, the skills were searched and extracted only from a subset of phrases with a high probability of containing at least one soft skill, namely the ones in which a clue appears.

Training The aim of this phase is to train a supervised model that is able to automatically extract soft skills from text. The outputs of the two previous phases helped to annotate the corpus required for the training of the model. In fact, as stated before, a supervised method requires an entity annotated corpus in order to extract new entities from unseen documents. We decided to train a classifier on the corpus of scientific publications used in the skill extraction phase. This corpus is suitable for two reasons. First, it contains a large number of skills; this helps to have more examples to deal with. Second, it is made of a lexicon that covers a wide range of skills. We annotated the document set with the information of the extraction context. The entity annotation schema for a single token is defined using a widely accepted BIO annotation scheme (Ramshaw et al. 1999):

- B-EXTR: the token is the beginning of an entity representing an extraction context;
- I-EXTR: the token is the continuation of a sequence of tokens representing an extraction context;
- O: for all the other cases.

In practice, it is specified in the corpus whenever a soft skill appears either alone or together with a clue. It is taken into account that an extraction context can appear with or without a clue. For example, the sentence "the evaluation of participants is based on the assessment of their level of critical thinking and problem solving" would be tagged as follow "the evaluation of participant is based on the assessment of their ¡extr¿ level of critical thinking¡/extr¿ and ¡extr¿ problem solving¡/extr¿". In our experiments the classifier has been trained using two different learning algorithms: Support Vector Machines (SVM) using the LIBSVM library (Chang et al, 2011) configured to use a linear kernel and Multi Layer Perceptron (MLP) implemented using the Keras library (Chollet 2015). We chose the SVM and MLP method to study how two wheel established (but efficient) state of the art classifiers perform on the specific task of soft skill extraction and to evaluate their performance in terms of precision and recall. The classifier uses the three features described in Section 3.3.2: lexical, grammatical and syntactic. We use the built-up parser of spacy with the "en_core_web_sm[20]" model that is an english multi-task CNN trained on OntoNotes. This model assigns context-specific POS tags and dependency parses using respectively the Universal POS tags and Universal Dependency Relations.

[20]https://spacy.io/models/enen$_{c}ore_{w}eb_{s}m$

Soft Skill	Source	Source Citations
Problem Solving	Hmelo-Silver et al, 2007	814
Active Reasoning		
Communication	Andrews, 2008	225
Professionalism		
Leadership	.Robles, 2012	218
Teamwork		
Flexibility		

Table 11: The seed list of soft skills, their source and the number of citations of the paper

3.4 Results

In this section we show the results of the application of the SkillNER in an experimental setup. In particular in section 3.4.1 we describe the results of the Clue Extraction, in section 3.4.2 the results of the Skill Extraction, in section 3.4.3 we show the result and evaluation of the trained models and finally in section 3.4.4 we validate the tool on a concurrent analysis in order to further proving its robustness.

3.4.1 Clue Extraction

The first phase of our methodology deals with the extraction of a list of clues (contextual words indicating the presence of a soft skill) . In order to collect valid clues, we used a seed list of soft skills, defined by using the soft skills mentioned in the most three cited papers on Scopus regarding soft skills as shown in table 11.

Among the skills declared as soft in these three papers, we kept in our seed list the ones already present in O*NET. For example, in Robles (2012) the “executive perceptions of the top 10 soft skills needed in today’s workplace” is discussed. Among the 10 skills listed by the author, we kept only the three (“leadership”, “teamwork”, “flexibility”) because also present in O*NET. This validation step is done in order to ensure to consider only reliable entries as seed list of soft skills. We perform the extraction of clues downloading 15 open-access documents from the Skills Panorama dataset provided by CEDEFOP[21]. We collected reports regarding replacement demand, skills challenges in Europe, polarisation of skills and job growth creators in the European labour market.

Table 12 shows the list of the most reliable clues extracted from this phase. It contains the top 5 clues based on their cumulative value of the different declinations of the same expression (such as “ability to”, “ability in”, “ability of”). The frequency value counts the number of times the clue appears before the one of the skills of the

[21]https://skillspanorama.cedefop.europa.eu/en

CLUE	Frequency
Ability (to/in/of)	64
Capability (to/in/of)	35
Level (in/of)	20
Know-how (in/of)	19
Proficiency (in/of/at)	10

Table 12: Clues List

seed list. As stated before, the higher is the frequency, the higher is the probability that the clue introduces a soft skill.

3.4.2 Skill Extraction

The clues defined in the previous phase are then used to search for skills. Figure 19 shows an example of how a clue could help us to find a skill. Once detected the skill "solve difficult problems" (skill contained in the seed list) we extract the clue "ability to" from the sentence. Then, searching for this clue in another sentence gave us the possibility to find the entity "explore the possibility of acquiring knowledge". This new entity belongs to our list of skills.

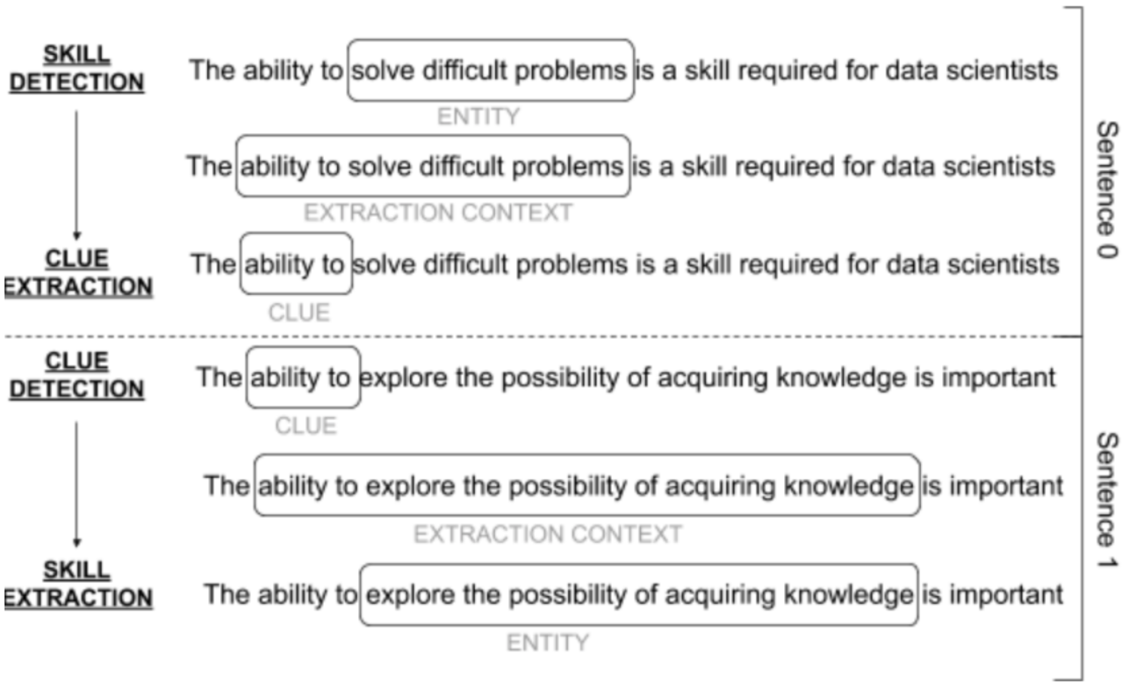


Figure 19: Example of a new entity extraction

Skills
Empathy
Abstract reasoning
Address emotions
Assertiveness
Compassion
Conflict management
Encode emotions
Empowering
Non verbal decoding
Manage a good diction
...

Table 13: Extraction of Soft Skill list

First, we collect all the phrases that contain a clue. Then, we extract the soft skills present in the collected sentences. This task has been manually executed thanks to a panel of domain experts with knowledge and expertise in the field of soft skills made of 4 psychologists and 4 Human Resources experts. The corpus that we use to perform these tasks is made of 5.359 abstracts of scientific papers in which the word "soft skill" occurs in the title, in the abstract and in the keywords. We identify 850 sentences containing at least one clue. Among these sentences we detected a list of 842 unique soft skills. These skills appear in the corpus with an overall frequency of 2.045. Figure 20 shows an example of a sentence containing a clue and three different soft skills, while table 13 shows a sample of the soft skills contained in the final list.

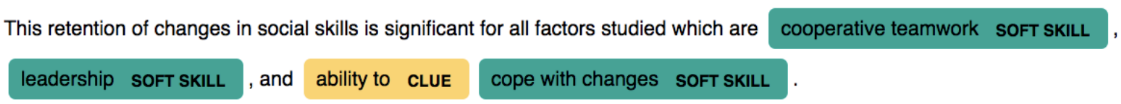


Figure 20: Example of sentence containing a clue

3.4.3 Experimental results and evaluation for the trained model

To train our model we use the same dataset for the previous phase, namely a corpus of 5.359 abstracts of scientific papers. We automatically annotated each abstract by projecting the input list of clues and soft skills. As a result we have a corpus in which the extraction context is annotated. After this task, the document set is splitted in training and test sets. To build an informative training set, from the corpus we select a subset of sentences containing at least one clue. The size of the training set in both cases is approximately composed by 1.800.623 tokens and 103.547 sentences. To test the performances of the implemented soft skill extractor, we devise two different configurations, each on using a specific learning algorithm. The main purpose of this procedure is to find the configurations that better perform the extraction task. We evaluated the NER system with the well-established measures of precision, recall and f1-score at a token level. Precision is the percentage of named entities found by the learning system that are correct; recall is the percentage of named entities present in the corpus that are found by the system; f1-score is the harmonic mean of the precision and recall. Table 14 reports the values of the defined metrics across the two experiments run with the two learning algorithms.

	Precision	**Recall**	**F1-score**
SVM	68.1	77.8	72.6
MLP	59.1	65.7	62.2

Table 14: Evaluation metrics

First, we can see that in both experiments the recall is higher than the precision. Second, we can see how the SVM-based training produces more reliable results. In fact, the F1-score is higher in the configuration of the Support Vector Machine. Based on the outcomes of the experiments, we use the first model for our SkillNER. A demonstration of the SkillNER is made available as a web-application[22]. We deploy the soft skill extraction model into a spacy model and we build an application with streamlit[23] where it is possible to try the SkillNER and to download the soft skills list. SkillNER can be used in two modes: rule-based mode and supervised mode. The first implies the simple search in the text of skills belonging to the skill list. In practice, it detects and extracts only pieces of text that have an exact match with the items of the list. However, the system is not able to detect a skill that is not present in our skill set. The second mode uses the model trained with the SVM. This implies that it not only performs the exact match of the tokens but it could also be able to detect new expressions of soft skills.

3.4.4 Concurrent Evaluation

The key elements of our tool are the list of skills and the list of clues. We populate these lists by using an approach that relies on the knowledge and expertise of a panel of experts that analyzed documents belonging to the state-of-the-art of this field. The annotation of the corpus has been done using the skills and clues that we defined in this search for soft skills. Therefore, it is necessary to further prove reliability of the whole system. We decided to perform a concurrent analysis using

[22]https://mysterious-hollows-20657.herokuapp.com/
[23]https://www.streamlit.io/

as input for the system the transversal skills defined by ESCO, assuming that the transversal skills of ESCO can be considered as a golden set of soft skills. We want to measure the differences in terms of performances between this system and the SkillNER. We train an entity extraction system using a SVM on a corpus annotated with the transversal skills using the same annotation schema described in Section 3.3.3. However, transversality is a complex concept. The skills mentioned in this dataset comprise those ones that explicitly refer to technical and digital skills. For this reason, we do not consider all the sub-categories which contain a reference to ICT and digital tools. We annotated the same 5.359 scientific papers described in the previous experiment, and then we split it into training and test set. We precision, recall and f1-score to evaluate the system. We compare the two systems based on their evaluation metrics shown in Table 15.

	Precision	**Recall**	**F1-score**
SkillNER	68.1	77.8	72.6
SVM on Transversal Skills	38.9	42.6	40.7

Table 15: Concurrent analysis

First, all the metrics show a better performance of the SkillNER. The low precision and recall proves that the automatic extraction process of the soft skills requires a system that takes into account not only a golden list of entities, but also requires contextual information. Our system is provided with such information. In fact, table 15 shows that using an approach that looks at the clues surrounding a skill helps to automatically extract them. Second, we also need to look at the data given as input to the two systems. The transversal skill list is made of 506 items, 336 less than the skills list used to train the SkillNER. The consistency of our skill list is reinforced by the fact that there is relevant overlap between them. In fact, 409 skills appear in both lists. This means that the 81% of transversal skills are already present in our skill list. Finally, this analysis proves that a NER system whose purpose is to extract entities whose definition is still not clear at the state of the art, works better if it has an input expressed in a very simple and concise way.

3.5 Case Study: Soft Skills and Job Profiles Clustering

3.5.1 Building and clustering relationships among Soft Skills and Job Profiles

As stated in section 3.1, it is still hard to define the relationships that exist between soft skills. The NER system developed in this study represents a first step in this direction, since it can be used by other scholars as a starting point for quantitative or qualitative studies. After having performed our concurrent evaluation, the results revealed us the opportunity to define what is currently missing on ESCO: the relationships among job profiles based on soft skills shared, and the relationships among soft skills based on job profiles in common. In this Section we demonstrate an application of our SkillNER:, we searched for the skills list in the overall ESCO database (preferred label, alternative label and description of skills). The sharing of a soft skill represents a relation between two professions, which may have various implications. The same reasoning could be applied to the dual situation, i.e. the share of job profiles among soft skills. This let the authors look at the phenomenon from two different perspectives, visualizing the relationships with two graph analyses. Even if ESCO is internationally recognized as a reliable source, the authors need to stress that the relationships that emerge from the clusterization of jobs and skills are strongly influenced by its structure; so, the present case study represents only a view on the topic, and could not be considered representative of the entire labor market.

Having said that, we use two approaches to measure respectively the relationships between soft skills and job profiles mentioned in ESCO:

- For the skill graph, the skills are joined by edges if they are mentioned in the same job profile and the edges between vertices have a weight that describes the strength of the relationship measured as a frequency value (total number of pairwise skill mentions);
- for the job graph, the jobs are joined by edges if they have in common a soft skill and the edges between vertices have a weight that describe the strength of the relationship measured as a frequency value (total number of pairwise skills in common).

To build the graph, we represent the counts as an (N, N) adjacency matrix, where N is the number of unique skills or job profiles and the elements in the matrix indicate the number of co-occurrences. We then use the adjacency matrices to generate an undirected graph for skill, Gs = (Vs, Es), where vertices, Vs, are the skills, and edges, Es, are the co-occurrences of the skills in the same job profile. We did the same for the job profiles graph Gj = (Vj,Ej) where vertices, Vj are the job profiles, and edges, Ej, are the number of skills the two vertices have in common. In other words, the higher the weight associated to the edge, the higher the co-occurrence of the skills or job profiles and the stronger the relationship between them.

3.5.2 Communities of Soft Skills and Job Profiles

To synthesize the information contained in the graphs, we grouped the skills and job profiles into clusters. In particular, we want to determine the intrinsic grouping in our set of entities (jobs or soft skills) in order to synthesize the results and characterise a small number of a representative set of entities. Regarding the clustering,

Graph	Nodes	Edges	Average in-degree
Gs: Skill graph	409	4336	8.54
Gj: Job graph	1243	23455	29.78

Table 16: Graph analysis results

since there is no absolute best criterion of choosing the right number of clusters, we chose to perform the Laplacian algorithm (Lambiotte et al. 2019) iteratively, changing the number of clusters as input. By evaluating the two parameters of modularity and resolution, that give information on how good the partition created is, we chose the proper number of clusters. Table 16 shows the results of the graph analysis for graph Gs and Gj. The job graph is more populated, since it has ten times the number of nodes and six times the number of edges of the skill graph. Job graph also has a higher average in-degree, which is a signal of the thickness of the relationships among nodes.

In Figure 21 and Figure 22 we show a representation of the two graphs using Gephi software (Bastian et al, 2009) with the Force Atlas algorithm (Jacomy et al, 2014) for the layout. With the layout used, two nodes in the graph are represented closely if they share an edge, and the closeness is proportional to edge weight. In this way also nodes that belong to the same communities of nodes (nodes that can be grouped into sets such that each set is densely connected internally) but do not share any edge, are represented closely. In other words, the visualizations tend to be coherent with the clustering algorithm. The size of the node is proportional to its in-degree while the colour expresses the cluster to which each node belongs. Finally, only the labels that are associated with nodes with higher frequency are shown. The process resulted in 20 communities for the Soft Skills graph and 14 communities for the Job Profile graph.

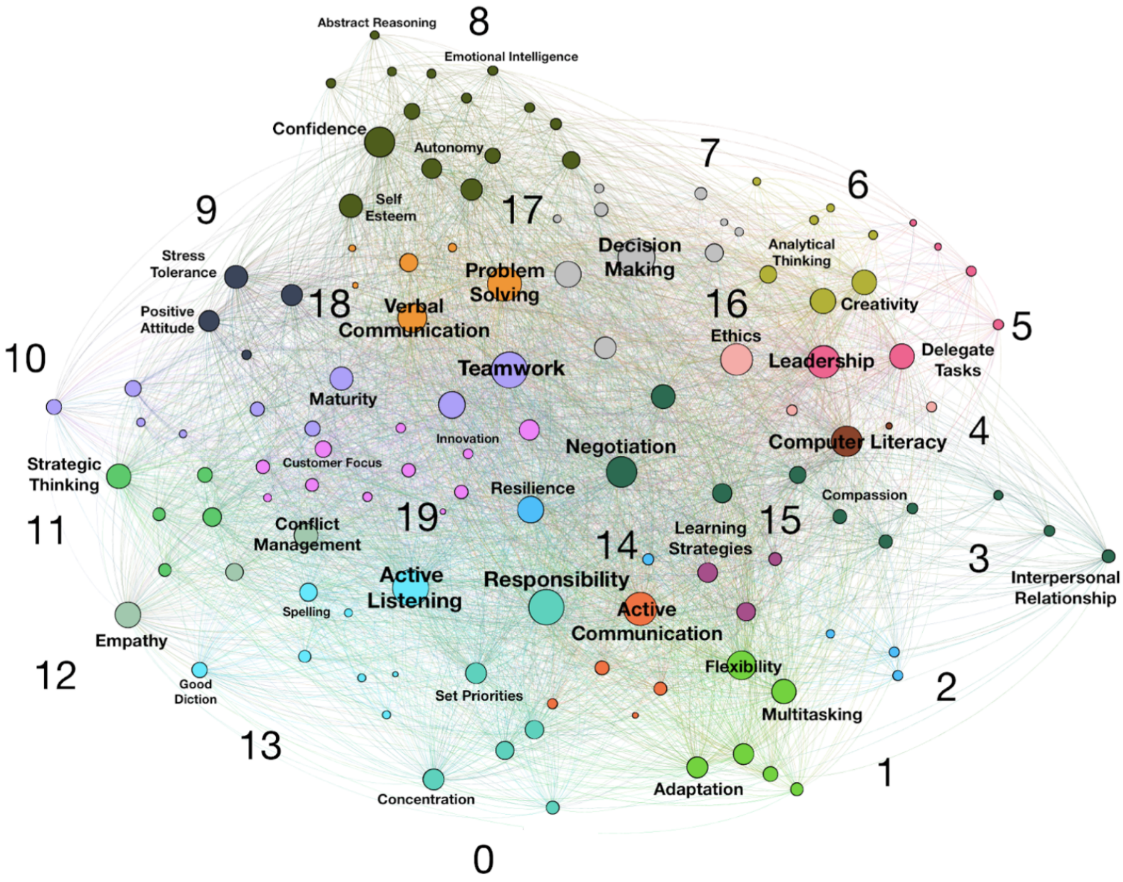


Figure 21: A network graph representation of the soft skills extracted and the clusters in which they are found. A soft skill is connected to another one if they have at least one job profile in common. Only the most relevant nodes are shown.

For what concerns the graph of soft skills (Figure 21), it shows mixed results: some of them were expected (for example, “leadership” and “delegate tasks” that share the same cluster), while others are surprising and they represent the true value of this visualization (i.e analytical thinking and creativity). In greater detail, the cluster number 6 demonstrates that there is a large number of jobs that possess the previous two skills jointly. It is also interesting to analyze cluster number 18 which shows the ability to solve problems strictly connected to communication; the previous insight has several implications, which deserve to be looked at from different perspectives and will be later analyzed. We should also highlight the close connection between empathy and conflict management; the communication cluster, which both includes active listening and good diction; finally, the ability of being multitasking, that is strictly associated with adaptability and flexibility.

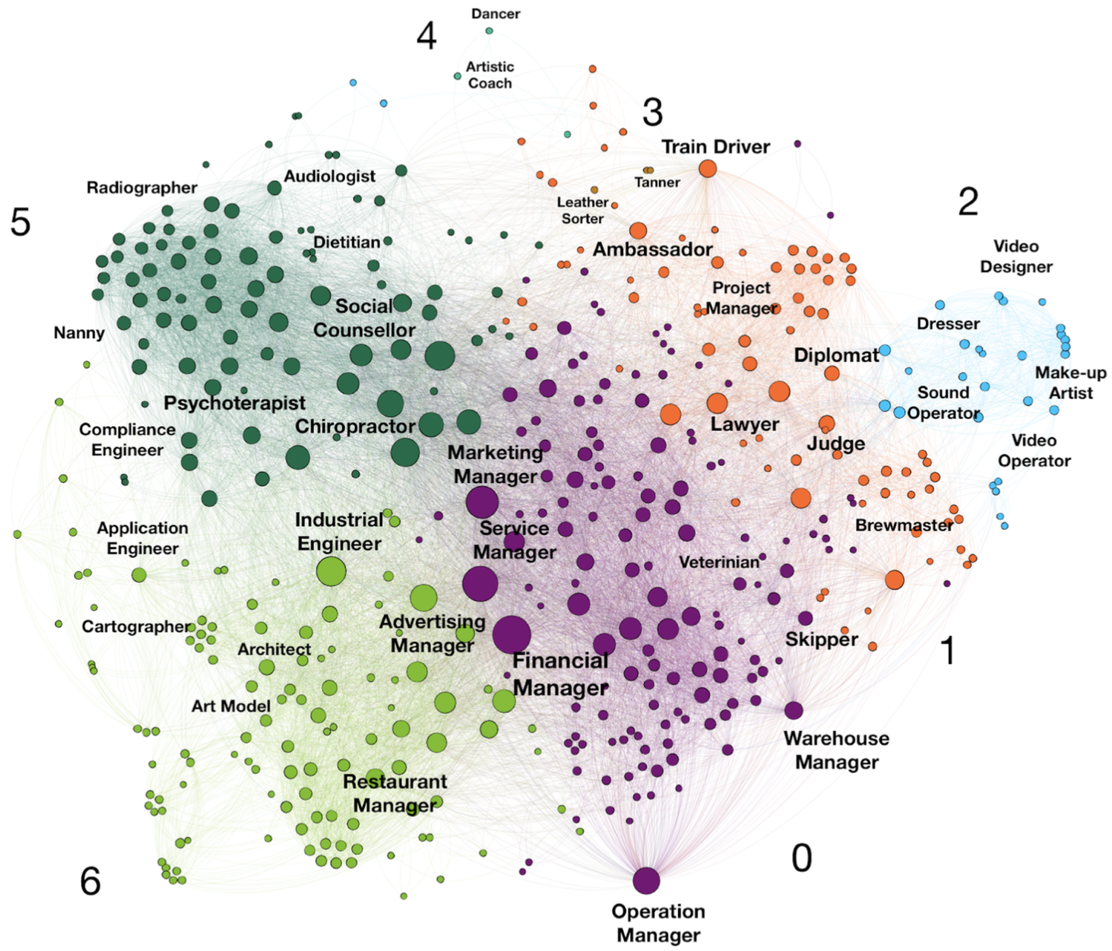


Figure 22: A network graph representation of the job profiles and the clusters in which they are found. A job profile is connected to another one if they have at least one soft skill in common.

The network graph displayed in Fig. 22 is made of 7 different clusters. Each node is a job profile and the size of the node gives and indication of the most soft-skilled occupations. A single cluster consists of workers which share similar soft skills. A good example is represented by cluster 2, which is made of artistic professions (i.e. Video designer, Make-up artists, Video-operator) that have in common soft skills such as creativity and originality. Another interesting cluster is number 0 which is populated by the biggest nodes, typically embodied by managerial roles (Marketing manager, Financial manager, Service manager). A greater size of the node implies a higher degree of the node itself, thus related jobs profiles are the ones that share more soft skills with other occupations. This happens for managers since they are professional figures with many soft skills in common, despite the sector in which they work (e.g. marketing, service, finance).

3.6 Discussion and research value

Output	Stakeholder	Purpose	Activity
Skill(N)ER	Firms	Recruitment	Skimming CV automatically
		Assessment	Identifying a lack of soft skills Developing ad hoc learning paths that ensure the portability of the acquired skills
	Institutions	Updating International Skills Databases	Laying the foundations for building a Soft Skills Ontology
Communities of Soft Skills	Firms	Updating Job Descriptions	Increasing Worker's Soft Skills, according to the relations that emerge from the graph
	Institutions	Updating International Skills Databases	
	Workers	Skill Portability	Identifying the skills more easily acquirable in relation to the ones currently owned
Communities of Job Profiles	Firms	Job-Knowledge Bridge Foresight Policy Design	Developing common Soft Skills courses for different job profiles, according to the relations visualized through the graph
	Institutions		Customizing University courses Making evidence of the most resilient job profiles
	Workers	Job Portability	Identifying the nearest job profile in relation to the one currently performed

Table 17: A summary table which reports the three outputs of the research, the stakeholders who may be interested about them, the final goals of the outputs and the activities that could be performed with them.

In the present section we summarise and describe the outputs of the work: the Skill(N)ER, the skills graph and job profiles graph. The Skill(N)ER applied to the adopted data sources extracted 850 phrases containing soft skills. The algorithm can support firms, making HR processes more precise, fast and reliable. In particular, the process of CV skimming for soft skills could be automated. Since the traceability to O*NET/ESCO skills are ensured by design, the tool allows the development of learning paths that also guarantee the portability of the acquired soft skills. Also Institutions may be supported, since the algorithm could help to update Internationally recognized databases and to further develop an ontology of soft skills (which seems to be missing at the moment). The skills graph represents an interesting output for several stakeholders. The graph visualizes the relationships among soft skills, an interesting added value since research is missing in the definition of their correlation (Hendon et al., 2017). In view of the above, firms and Institutions could use it as a guideline to enrich both skills' taxonomies and job descriptions, since it represents a reliable source of knowledge. The authors should also outline that every cluster could be seen as a prospective course program for both the stakeholders. Moreover, the closeness between clusters may suggest the right way to build effective learning paths. With respect to workers, we can introduce the concept of portability, defined in computing as the ability of a software to be transferred from a machine to another. In the labor market context, portability is the capacity to easily move from a profession to another thanks to the skills owned by the worker. The skills graph gives a visual representation of their proximity and the underlying learning paths; thus, it also provides evidence of the easiness of acquisition of a skill in relation to the ones already possessed. For what concerns the job profiles graph, the greater the size of the node, the more the profile has soft skills. Moreover, the greater the difficulty to decode a skill, the less it can be replicated by a machine (Acemoglu et al., 2010). Thus, bigger nodes in size could represent profiles potentially resilient to the Industry 4.0 phenomenon. In addition, the relation between the two graphs could be used as a support to the customization of University courses and to create the desirable bridge between Jobs and Knowledge. Another critical aspect is represented by the interconnections among jobs which are conventionally considered very distant from each other, thus giving companies the opportunity to develop soft skills courses to bring together the different occupations (for example, Architect and Industrial Engineer). Finally, the graph could be used by workers to identify their nearest job profile in relation to the profession currently performed, reinventing himself/herself in such an era of labor uncertainty. To summarise, we took into consideration the need for different potential stakeholders and the characteristics of different documents where the tool can be applied. In section 3.6.1 we list the potential stakeholders of the methods and in section 3.6.2 the potential fields of application of the system.

3.6.1 Potential Users of the System

The outputs of the research (the soft skills list, the graph of soft skills and the graph of job profiles) could bring multiple benefits to heterogeneous stakeholders.

- Companies: to design effective training courses, to update job descriptions and to improve the assessment of internal soft skills;
- Policy Makers: to get an early indication about the likely resilience of job profiles and to design policies accordingly;
- Researchers: to identify the most resilient job profiles and to facilitate the foresight of labor market changes;
- Recruiters: to make more reliable and efficient the recruiting process;
- Universities: to create new curricula and courses, facilitating the job-knowledge bridge;

- EU Agencies: to lay the foundations for building an ontology of soft skills but also to update the current skill taxonomies;
- Unemployers: to identify the nearest job profile (and acquirable skills) according to the one currently performed (and possessed).

3.6.2 Fields of Application

- Occupational Frameworks: to identify the standard label of a skill and to assign it all its periphrases, creating a link between the standard and the different ways it is actually expressed;
- Job Vacancies: to identify the soft skills mostly required by labor market;
- Curricula: to identify the most offered soft skills, making also a comparison between offer and demand;
- Job Descriptions: to homogenise the different ways the same skill is expressed in a single company;
- Scientific Papers: to represent the trend of interest of specific soft skills and make a comparison among them;
- MOOCs (Massive Online Open Courses): to identify the sub-skills of a specific soft skill, studying the content of modules, which would additionally allow the design of customized learning paths;
- Social Networks: to detect the most hype soft skills but also the alternative (and mostly colloquial) ways in which they are expressed.

3.7 Conclusions and Future Developments

The importance of soft skills is striking in a world that pushes steadily more and more towards digitalization. The main reasons can be found within the complexity of labor market dynamics and economy, the advent of the fourth industrial revolution and its rate of change. As (Acemoglu et al., 2010; Levy et al., 2004) stated, machines can replicate what is easily encoded. Furthermore, the lack of a universal definition of soft skills domain in contemporary literature is a proof of a low percentage of their current encoding (and a lack of their overall comprehension as well) , which makes their ownership ever more strategic. On those grounds, the aim of this paper is to develop an automatic method based on NLP techniques, in order to collect a list of soft skills from heterogeneous sources. The final goal is trying to delineate the boundaries of this fuzzy domain, developing a scalable and fast tool. This work helped us to define a soft skills list using reliable sources and then to build a NER system able to identify soft skills in any text. Besides to increase reproducibility in this field of study, we produced an open-source database to be used by other scholars; then we developed a tool hereafter easily adaptable to other domains. After that, we demonstrated the reliability of our extractions, the model adopted and the quality of the list extracted through the aid of a skill golden standard: ESCO. Once validated, we showed an application of the list in a case study on the same database. From this data we built a graph-based representation of the relations among soft skills and job profiles, providing a map of soft skills and jobs that can help the scientific community to give a clearer definition of what soft skills are. For what concerns future developments, we would additionally improve the methodology by using a mixed deep learning and rule based approach. Using new language representation models such as BERT (Devlin et al., 2018), is nowadays possible to have a more effective representation of words as meaningful vectors. It is possible to integrate this system with rules based systems, increasing thus both recall and precision of our system. Another further step can be done applying the algorithm on other HR related data sources (Job Descriptions, Curricula), technical sources (patents), open dictionaries (wikipedia) or social networks (twitter). This can make it possible to test the method on different domains, validating the tool and expanding the list of soft skills. Finally, we think that the results could make it easier for other scholars to define a shared list of soft skills, in order to improve the quality of the research in this field.

3.8 References

Abujabal, A., Roy, R., Weikum, G., (2018). *Never-ending learning for open-domain question answering over knowledge bases.* 1053-1062. 10.1145/3178876.3186004.

Acemoglu, D., Autor, D., (2010). *Skills, tasks and technologies: implications for employment and earnings.* doi:10.3386/w16082

Alabdulkareem, A., Frank, M. R., Sun, L., Alshebli, B., Hidalgo, C., Rahwan, I., (2018). *Unpacking the polarization of workplace skills.* Science Advances,4(7). doi:10.1126/sciadv.aao6030

Alfonso-Hermelo, D., Langlais, P., Bourg, L., (2019). *Automatically learning a human-resource ontology from professional Social-Network Data.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11489 LNAI, pp. 132-145. DOI: 10.1007/978-3-030-18305-9_11

Al-Nabki, M.W., Fidalgo, E., Alegre, E., Fernández-Robles, L., (2019). *ToRank: identifying the most influential suspicious domains in the Tor network.* Expert Systems with Applications, 123, pp. 212-226. DOI: 10.1016/j.eswa.2019.01.029

Amal, S., Tsai, C.-H., Brusilovsky, P., Kuflik, T., Minkov, E., (2019). *Relational social recommendation: application to the academic domain.* Expert Systems with Applications, 124, pp. 182-195. DOI: 10.1016/j.eswa.2019.01.061

Andrews, J., Higson, H., (2008). *Graduate employability, 'soft skills' versus 'hard' business knowledge: A european study.* Higher Education in Europe, 33 (4), pp. 411-422. Cited 225 times.

Autor, D., Dorn, D., (2009). *The growth of low skill service jobs and the polarization of the U.S. Labor Market.* doi:10.3386/w15150

Bauer, C., Viola, K., Strauss, C., (2011). *Management skills for artists: 'learning by doing'?* Int. J. Cult. Policy 17 (5), 626–644.

Bastian M., Heymann S., Jacomy M. (2009). *Gephi: an open source software for exploring and manipulating networks.* International AAAI Conference on Weblogs and Social Media.

Blake, R., Gutierrez, O. (2011). *A semantic analysis approach for assessing* *professionalism using free-form text entered online.* Computers in Human Behavior, 27(6), 2249–2262. doi: 10.1016/j.chb.2011.07.004

Blanco-Fernández, Y., Gil-Solla, A., Pazos-Arias, J.J., Ramos-Cabrer, M., Daif, A., López-Nores, M.,(2020). *Distracting users as per their knowledge: combining linked open data and word embeddings to enhance history learning* . Expert Systems with Applications, 143, art. no. 113051. DOI: 10.1016/j.eswa.2019.113051

Blondel, V., Jean-Loup, G., Lambiotte R., Lefebvre E., (2008). *Fast unfolding of communities in large networks.* Journal of Statistical Mechanics: Theory and Experiment (10), P1000

Bohlouli, M., Mittas, N., Kakarontzas, G., Theodosiou, T., Angelis, L., Fathi, M., (2017). *Competence assessment as an expert system for human resource management: A mathematical approach.* Expert Systems with Applications, 70, pp. 83-102. DOI: 10.1016/j.eswa.2016.10.046

Boselli, R., Cesarini, M., Mercorio, F., Mezzanzanica, M., (2018). *Classifying online job advertisements through machine learning.* Future Generation Computer Systems,86, 319-328. doi:10.1016/j.future.2018.03.035

Bridgstock, R., (2011). *Skills for creative industries graduate success.* Educt. Training 53 (1), 9–26.

Brynjolfsson, E., McAfee, A. (2014). *The second machine age: work, progress, and prosperity in a time of brilliant technologies* . New York: W.W. Norton Company.

Cacciolatti, L., et al., (2017). *Clashing institutional interests in skills between government and industry: An analysis of demand for technical and soft skills of Graduates in the UK.* Technological Forecasting and Social Change

Chang C.-C., Lin C.-J., (2011). *LIBSVM: a library for support vector machines.* ACM Transactions on Intelligent Systems and Technology, 2. 27:1–27:27, software available at http://www.csie.ntu.edu.tw/cjlin/libsvm

Chechurin, L., Borgianni, Y., (2016). *Understanding TRIZ through the review of top cited publications.* Computers in Industry, 82, pp. 119-134.

Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G. (2018). *Extracting and mapping industry 4.0 technologies using Wikipedia.* Computers in Industry. 100.

244-257.

Chinchor N., Robinson P., (1997). *Muc-7 named entity task definition*, Proceedings of the 7th Conference on Message Understanding, p. 29.

Chollet,F., Keras,(2015) https://github.com/fchollet/keras.

Choudhary, A.K., Oluikpe, P.I., Harding, J.A., Carrillo, P.M., (2009). *The needs and benefits of text mining applications on ost-project reviews.* Computers in Industry, 60 (9), pp. 728-740.

Chryssolouris, G., Mavrikios, D., Mourtzis, D., (2013). *Manufacturing systems: skills competencies for the future.* Procedia CIRP 7, 17–24.

Ciaramita M., Altun Y., (2005). *Named-entity recognition in novel domains with external lexical knowledge,*, Proceedings of the NIPS Workshop on Advances in Structured Learning for Text and Speech Processing, Whistler, British Columbia.

Cinque, M., (2015). Comparative analysis on the state of the art of Soft Skill identification and training in Europe and some Third Countries. Speech at "Soft Skills and their role in employability – New perspectives in teaching, assessment and certification", workshop in Bertinoro, FC, Italy. cle 28). ELT Newsletter.

Colombo, E., Mercorio, F., Mezzanzanica, M., (2019). AI meets labor market: Exploring the link between automation and skills. Information Economics and Policy, 47, pp. 27-37. DOI: 10.1016/j.infoecopol.2019.05.003

Cooper, R., Tang, T., 2010. *The attributes for career success in the mass communication industries: a comparison of current and aspiring professionals.* J. Mass Commun. Educ. 65 (1), 40–55 (Mar.).

Deming D. Kahn L., (2018). *Skill requirements across firms and labor markets: evidence from job postings for professionals.* Journal of Labor Economics, vol 36(S1), pages S337-S369.

Devlin, J., Chang, M., Lee, K., Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding.*

Dobrunz, J., Schöppner, H., Wolfram, A., (2006). *Anforderungen an den Mediennachwuchs* — die media.net employability-studie. media.net berlinbrandenburg e.V, Berlin.

Duran-Novoa, R., Leon-Rovira, N., Aguayo-Tellez, H., Said, D., (2011). *Inventive problem solving based on dialectical negation, using evolutionary algorithms and TRIZ heuristics.* Computers in Industry, 62 (4), pp. 437-445.

Etzioni O., Cafarella M., Downey D., Popescu A.-M., Shaked T., Soderland S., Weld D. S., Yates A., (2005). *Unsupervised named entity extraction from the web: An experimental study.* Artificial intelligence, vol. 165, no. 1, pp. 91–134.

Evenson, R.,(1999). *Soft skills, hard sell.* Techniques: Making Education and Career Connections, v74 n3 p29-31

Fernández, N., Arias Fisteus, J., Sánchez, L., López, G. (2012). *IdentityRank: Named entity disambiguation in the news domain.* Expert Systems with Applications, 39 (10), pp. 9207-9221. DOI: 10.1016/j.eswa.2012.02.084

Fernández-Sanz, L., Gómez-Pérez, J., Castillo-Martínez, A., (2017). *E-Skills Match: A framework for mapping and integrating the main skills, knowledge and competence standards and models for ICT occupations.* Computer Standards and Interfaces, 51, pp. 30-42. DOI: 10.1016/j.csi.2016.11.004

Fernbach S., Sloman P., (2017). *The Knowledge Illusion: why we never think alone.* Riverhead Books

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Rahwan, I. (2019). *Toward understanding the impact of artificial intelligence on labor.* Proceedings of the National Academy of Sciences, 116(14), 6531–6539. doi: 10.1073/pnas.1900949116

Frey C.B., Osborne M.A., (2017). *The future of employment: How susceptible are jobs to computerization?* Technological Forecasting and Social Change, Sept 1–72.

Galati, F., Bigliardi, B., (2019). *Industry 4.0: Emerging themes and future research avenues using a text mining approach.* Computers in Industry, 109, pp. 100-113.

Gildea D., (2001). *Corpus variation and parser performance,* Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing, pp. 167–202.

Gorecky, D., Schmitt, M., Loskyll, M., Zühlke, D., (2014). *Human-machine-interaction in the Industry 4.0 era.* In: In Proceedings of the 2014 12th IEEE International Conference on Industrial Informatics (INDIN), Porto Alegre, Brazil.

Hanskel J., Westlake S., (2018). *Capitalism without Capital: The Rise of the Intangible Economy.* Princeton University Press

Harrag, F. (2014). *Text mining approach for knowledge extraction in Sahîh Al-Bukhari.* Computers in Human Behavior. 30. 558-566. 10.1016/j.chb.2013.06.035.

Haukka, S., (2011). *Education-to-work transitions of aspiring creatives.* Cultural Trends 20 (1), 41–64.

Hendon, M., Powell, L., Wimmer, H. (2017). *Emotional intelligence and communication levels in information technology professionals.* Computers in Human Behavior, 71, 165–171. doi: 10.1016/j.chb.2017.01.048

Hmelo-Silver, C. E., Duncan, R. G., Chinn, C. A. (2007). *Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner*, Sweller, and Clark. Educational Psychologist,42(2), 99-107.

Hong L., Davison B.D., (2010). *Empirical study of topic modeling in Twitter.* Dept. of Computer Science and Engineering.

Honnibal, M. and Montani, I. (2017). *Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.* https://spacy.io/usage/rule-based-matching

Jacomy, M., Venturini, T., Heymann, S., Bastian, M. (2014). *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi Software.* PloS one. 9. e98679. 10.1371/journal.pone.0098679.

James, R. F., James, M. L. (2004). *Teaching career and technical skills in a "mini" business world.* Business Education Forum, 59(2), 39-41.

Karakatsanis, I., Alkhader, W., Maccrory, F., Alibasic, A., Omar, M. A., Aung, Z., Woon, W. L., (2017). *Data mining approach to monitoring the requirements of the job market: A case study.* Information Systems,65, 1-6. doi:10.1016/j.is.2016.10.009

Kemper, C.L. (1999). *EQ vs. IQ.* Communication World,16(9), 15-19.

Krasnashchok, K., Jouili, S. (2018). *Improving topic quality by promoting named entities in topic modeling.* 247-253. 10.18653/v1/P18-2040.

Lafferty J., McCallum A., Pereira F., (2001). *Conditional random fields: Probabilistic models for segmenting and labeling sequence data.* In: Proceedings of the eighteenth international conference on machine learning, ICML. vol. 1; p. 282–289.

Lambiotte R., Delvenne J.-C., Barahona M., (2009). *Laplacian dynamics and multiscale modular structure in networks,* arXiv. 1.

Last, C., (2017). *Global commons in the global brain.* Technological Forecasting and Social Change, 114:48–64.

Levy, F., Murnane, R. (2004). *The new division of labor: how computers are creating the next job market.* New York; Princeton; Oxford: Princeton University Press.

Liu, L., Li, Y., Xiong, Y., Cavallucci, D., (2020). *A new function-based patent knowledge retrieval tool for conceptual design of innovative products.* Computers in Industry, 115, art. no. 103154.

Lo, S.L., Chiong, R., Cornforth, D., (2017). *An unsupervised multilingual approach for online social media topic identification.* Expert Systems with Applications, 81, pp. 282-298. DOI: 10.1016/j.eswa.2017.03.029

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.

McCallum A., Freitag D., Pereira FC., (2000). *Maximum entropy Markov models for information extraction and segmentation.* In: Icml. vol. 17; p. 591–598.

McMurchie, L.L. (1998). *Careers can rise or fall with*
EQ.Computing Canada, 1(9), 18-21.

Mirski, P., Bernsteiner, R., Radi, D., (2017). *Analytics in human resource management: The OpenSKIMR approach.* Procedia Computer Science,122, 727-734. doi:10.1016/j.procs.2017.11.430

Mitchell, Geana Skinner, Leane White, Bonnie, (2010). *Essential soft skills for success in the Twenty-first Century: workforce as perceived by business educators.* Delta Pi Epsilon Journal. 52. 43-53.

Nadeau D., Sekine S., (2007). *A survey of named entity recognition and classification.* Lingvisticae Investigationes, vol. 30, no. 1, pp. 3–26,

Nieragden, G. (2000). *The soft skills of business English* (The Weekly Column: Arti-)

Nicoletti, M., Schiaffino, S., Godoy, D., (2013). *Mining interests for user profiling in electronic conversations.* Expert Systems with Applications, 40 (2), pp. 638-645. DOI: 10.1016/j.eswa.2012.07.075

Oliva, J.T., Lee, H.D., Spolaôr, N., Takaki, W.S.R., Coy, C.S.R., Fagundes, J.J., Wu, F.C.,(2019). *A computational system based on ontologies to automate the mapping process of medical reports into structured databases.* Expert Systems with Applications, 115, pp. 37-56. DOI: 10.1016/j.eswa.2018.08.004

Perreault, H. (2004). *Business educators can take a leadership role in character education.* Business Education Forum, 59, 23-24.

Piskorski J., Yangarber R., (2013). *Information extraction: past, present and future.* Multisource, Multilingual Information Extraction and Summarization, Springer, pp. 23–49.

Pryima, S., Rogushina, J.V., Strokan, V., (2018). *Use of semantic technologies in the process of recognizing the outcomes of non-formal and informal learning.* CEUR Workshop Proceedings, 2139, pp. 226-235.

Quinlan JR., (1986). *Induction of decision trees.* Machine learning. 1(1):81–106

Ramshaw L.A., Marcus M.P., (1999). *Text chunking using transformation-based learning.* Natural Language Processing Using Very Large Corpora, Springer, pp. 157-176

Ritter A., Clark S., O. Etzioni O., (2011). *Named entity recognition in tweets: an experimental study.* Proc. EMNLP, pp. 1524– 1534

Robles, M. M. (2012). *Executive perceptions of the Top 10 soft skills needed in today's workplace.* Business Communication Quarterly,75(4), 453-465.

Rodrigues, C.M.D.O., Freitas, F.L.G.D., Barreiros, E.F.S., Azevedo, R.R.D., de Almeida Filho, A.T., (2019). *Legal ontologies over time: A systematic mapping* *study.* Expert Systems with Applications, 130, pp. 12-30

Sanz, I., Montero, J., Sevillano, X., Carrié, J.C., (2019). *Developing a videogame for learning signal processing and project management using project-oriented learning in ICT engineering degrees.* Computers in Human Behavior. 99.

Sari, Y., Hassan, M.F., Zamin, N., (2010). *Rule-based pattern extractor and Named Entity Recognition: A hybrid approach.* Proceedings 2010 International Symposium on Information Technology - Engineering Technology, ITSim'10. 2. 563 - 568. 10.1109/ITSIM.2010.5561392.

Sarica, S., Luo, J., Wood, K.L., (2020). *TechNet: Technology semantic network based on patent data.* Expert Systems with Applications, 142, art. no. 112995. DOI: 10.1016/j.eswa.2019.112995

Schulz, B. (2008). *The importance of soft skills: Education beyond academic knowledge.* Journal of Language and Communication. 2. 10.1016/0006-3207(93)90452-7.

Sung, T. K. (2018). *Industry 4.0: A korea perspective.* Technological Forecasting and Social Change, 132:40–45.

Suykens JA, Vandewalle J., (1999). *Least squares support vector machine classifiers.* Neural processing letters, 9(3):293–300.

Szalavetz, A., (2018). *Industry 4.0 and capability development in manufacturing subsidiaries.* Technological Forecasting Social Change,
https://doi.org/10.1016/j.techfore.2018.06.027

Thompson, K. R., Koys, D. J. (2010). *The management curriculum and assessment journey: Use of Baldrige criteria and the occupational network database.* Journal of Leadership Organizational Studies, 17(2), 156–166.

Tseng, H., Yi, X., Yeh, H.-T. (2019). *Learning-related soft skills among online business students in higher education: Grade level and managerial role differences in self-regulation, motivation, and social skill.* Computers in Human Behavior, 95, 179–186. doi: 10.1016/j.chb.2018.11.035

Van Laar, E., Deursen, A., Van Dijk, J., Haan, J. (2017). *The relation between 21st-century skills and digital skills: A systematic literature review.* Computers in Human Behavior. 72. 10.1016/j.chb.2017.03.010.

Weber, E., (2016). *Industry 4.0: Job-producer or Employment-destroyer?* Retrieved on March 2017 from
http://doku.iab.de/aktuell/2016/aktueller_bericht_1602.pdf.

4 Who rises and who drops? a comparative analysis of labor demand changements in Emilia-Romagna[24]

4.1 Introduction

New technologies, new paradigms, new skills. An industrial revolution is a radical change in socio-economic and political systems, driven by the introduction of new technologies usually (but not necessarily) associated with a significant increase in efficiency and productivity (Tarry, 2019). From the beginning of the First Industrial Revolution, modern capitalism started to be based on knowledge economy, since the true value was produced, in large part, by the propagation of the uses of available knowledge. In other words, human work does not transform the raw material, but - if we refer to cognitive work - it generates innovative knowledge (Rullani, 2004). The latter is used to transform what already exists, and indirectly creating utility, increasing efficiency, designing new products or customizing old services (Rullani, 2004). For this reason, in an innovation-based economy, the interdependencies in terms of learning and innovations are definitely crucial (Grillitsch et al., 2018).

What was previously stated, is ever more relevant when we refer to Industry 4.0. Of the Fourth Industrial Revolution, also known as Industry 4.0, literature provides many different and sometimes discordant definitions, which derive from different interpretations. What they have in common is the idea that its impact cuts across the social, economic and political, with boundaries that are not yet clearly defined (Last, 2017).

As regards the effects of Industry 4.0 on the labour market, there are several schools of thought. On the one hand, automated systems, the backbone of the 4.0 paradigm, are considered to be a threat: some professions may be at risk of replacement by machines and algorithms. The ultimate reason is that artificial intelligence and big data are giving machines more and more human abilities (Rotman, 2013).

As regards the likelihood of workers being replaced by machines, the research of Frey et al. (2017) is emblematic: they estimated that around 47 % of jobs were in the high-risk category, in particular those with routine activities.

Conversely, Caruso (2017) states that technological innovation is not replacing work but is creating new opportunities through increased process efficiency. Similarly, Rosenberg (2009) sees automation not a threat but an opportunity: if workers were replaced by machines in routine and low-value activities, they could be more freer to express their talents. As a result, new technologies could have a positive impact on employment: in particular, 3D printing, the Internet of Things, augmented reality and big data analysis require significant numbers of new skills for proper manage-

[24] *Fareri, S., Silvestri, L., Caruso, G., Solinas, G. (2020). "Who rises and who drops? a comparative analysis of labor demand changements in Emilia-Romagna"*

ment (Freddi, 2017), leading to demand for new professions. The same school of thoughts belongs to (Leight et al, 2020) that highlight a positive contribution of Robots to manufacturing employment in USA. In the same vein, MacCrory (2014) assumes three main consequences of technological innovation: a significant reduction in skills in competition with automation; a significant increase in skills which complement machines; finally, an increase in the strategic nature of skills which cannot be replicated by machines. Similar conclusions are reached by (Autor, 2015). Many of the topics (and interpretative hypotheses) which have characterised the comparison on the effects of technical progress among social scientists since the First Industrial Revolution are now being brought up again, both in terms of employment levels and in relation to the conditions and quality of work.

Although within different visions, the speed of the current processes, their transversality and the very unpredictability of the directions of change – in a word the characteristic aspects of the Fourth Industrial Revolution – pose unavoidable questions about the expected evolution of labour demand, both in quantitative and qualitative terms. More than half a century ago, the pioneering work carried out by the United States Bureau of Labor Statistics (BLS) strongly emphasised the need to pay attention to the evolution of the demand for new skills associated with the transformation of technologies and production processes (Wilson, 2013). Today, the proactive adaptation that is induced by the processes of change also involves the need not only to predict the "new", but also to manage the "old" skills and to plan for retraining. From a managerial point of view, the presence of a wrong mix of skills determines poor business performance (Grugulis et al., 2009; Lorentz et al., 2013): human capital represents the most valuable asset for an organization (Fulmer et al., 2013), and for this reason firms needs to carefully focus on its valorization (Barney, 1991; Becker et al., 1996; Lado et al., 1994). Moreover, it is precisely because organisations are complex entities that identifying the right mix of skills for a business is a very difficult task (Abbott, 1993). However, what seems evident is that robotisation is growing faster than the capacity of workers to acquire new skills; for this reason, there is the urgent need to design learning path to prepare workers to exploit technologies the most (Gentili et al., 2020) but also to create new tasks where labour can be productively employed synergistically with technologies (Acemoglu et al., 2020). In view of this, Artificial Intelligence itself (and associated technologies) could be used to positively transform education (Clifton et al., 2020)

In this context, several researchers have set themselves the objective of defining the characteristics that could make a professional profile resilient to change (Chryssolouris et al., 2013; Gorecky et al., 2014; Weber, 2016). In recent years, scholars, managers and operators, starting out from the same questions, have shown growing interest in cross-cutting skills. The focus of scholars and professionals on this topic

has grown for many reasons. The main one is, perhaps, digitisation itself. If the impact of digitisation is pervasive, diversified and articulated across multiple levels of skills and capabilities (Van Laar et al, 2017; Galati et al., 2017), the importance of acquiring cross-cutting and heterogeneous skills to deal with it becomes evident (Chryssolouris et al., 2013; Gorecky et al., 2014; Weber, 2016; Fonseca et al., 2018; Lalé, 2020)).
Although there is no agreement on the domain in which soft skills operate, seminal essays on the impact of digitisation on skills (Acemoglu et al., 2010; Author, 2015; Frey and Osborne, 2017; Levy et al., 2004; de Vries et al., 2020) show that the machines and algorithms that govern them are able to replicate solely what is codifiable. Cross-cutting skills, in this area, are the real bottleneck of digitisation.

This paper starts out from the state of the art which has quickly developed. A key source of information is represented by the taxonomies of skills. Many of these can be found online, but the main frameworks are ESCO (European) and O*NET (American).
ESCO (European Skill/Competence Qualification and Occupation) is a multilingual classification system for Europe: it ranks skills, expertise and qualifications in Europe relevant to the labour market. ESCO's goal is to bridge the gap between universities and industry across Europe, through a triangular relationship between skills, profiles and qualifications. The classification code for professions is ISCO-O8, which is the International Standard Classification of Occupations (International Labour Organisation, 2008). O*NET is the American equivalent of ESCO. O*NET is a database consisting of 974 occupations of the Standard Occupational Classification (SOC) and their corresponding skills, knowledge and abilities. The use of SOC allows us to analyse professions from multiple perspectives, comparing data from different sources, aggregating and monitoring them over time. Each professional profile contains quantitative indications of the level of proficiency required and the relative importance of each of the skills. ESCO also has a higher level of detail than O*NET, with six times the number of skills and three times the number of professional profiles (compared to the US counterpart). Finally, there is no clear distinction between hard and soft skills in O*NET, while about 110 skills are labelled as transversal in ESCO (v1.0.3); however, ESCO also has some issues, for example because its cross-cutting skills are too abstract to be assigned to a specific professional profile and their number is still limited. The two databases are the starting point for the aggregate study of skills and professional profiles in complex economic systems (regions, nations, groups of enterprises, etc.). However, being a static representation (and strongly affected by the time required to translate observation of production and training processes into a shared classification system), they were only minimally able to capture a highly dynamic phenomenon such as the evolution of demand for skills and abilities in the labour market.

Nonetheless, many authors used International framework of skills as valuable source of information with different objectives and analytical methodologies. Some of them belong to the field of econometrics (Frey et al., 2017; Arntz et al., 2017; Autor et al., 2009; Acemoglu et al., 2010; MacCrory et al., 2014), trying to measure the susceptibility of workers to automation; others are mainly based on data mining techniques, focusing on the development of further skills taxonomies or entity extraction (Boselli et al., 2018; Karakatsanis et al., 2017; Alabdulkareem et al., 2018; Colombo et al., 2019), data comparison (Thompson et al., 2010) and data linking (Mirski et al., 2017; Fernandéz-Sanz et al., 2017; Alfonso-Hermelo et al., 2019, Pryima et al, 2018). For what concerns specifically ESCO, it is frequently used to improve recruitment process (Mirski et al., 2017; Alfonso-Hermelo et al., 2019; Pryima et al, 2018).

What seems interesting to deepen, and could represent a possible step forward, is linking the skills in the classification systems described above with the professions present in the labour market, which is considered through the connection between the ISCO08 international classification and the professional units of the 2011 Classification of Professions (ISTAT). This latter classification is the latest in a long tradition of studies, with which the Italian National Institute of Statistics has sought to meet the needs for renewal which have emerged from many sides, particularly the institutions which work most for and on the labour market. Since 1861, updates to the classification of professions have, indeed, followed the deadlines of the general population census – the last of which was held in 2011. This last update "has enabled the instrument to be withdrawn and its structure to take account of changes in the labour market. Innovation in production processes and their organisation, new aspects in the qualifications required to exercise professions, and changes in demand for goods and services are just some of the factors that affect the nature, content and manner in which the different occupations are carried out. As these changes gradually become apparent, the classification needs to be adapted to reflect market trends, new occupational areas and changes in the requirements associated with the professions. The work of reviewing the taxonomy has been directed towards this point of view, in an attempt to grasp the changes in the professional structure of the country and to represent them within the new classification system." (ISTAT, 2013).
For what concerns the concept of skill, and what the authors mean when they cite the term skill, they refer to ESCO and EQF definition: "skill means the ability to apply knowledge and use know-how to complete tasks and solve problems. They can be described as cognitive (involving the use of logical, intuitive and creative thinking) or practical (involving manual dexterity and the use of methods, materials, tools and instruments)" (European Commission, 2020). In addition, the authors are aligned with ESCO distinction between "skill" and "competence", as the term

skill refers typically to the use of methods or instruments in relation to defined tasks, while the term competence is broader and refers typically to the ability of a person - facing new situations and unforeseen challenges through the use and apply of knowledge and skills in an independent way (European Commission, 2020).
The ability to link the classification systems for professions and skills has enabled the labour market administrative data (when accompanied by the ISTAT profession) to be combined with the skills associated with them and then examine the development of the demand for skills both retrospectively and prospectively. Even if literature shows several researches analysing Excelsior Informative System[25] as primary source of labor market data (Antonelli et al., 2008; Batazzi, 2015; Autiero et al., 2020), in the present work the authors propose SILER as the main one. SILER is a dataset of mandatory notifications sent to regional employment centres, made by more than three million observations, characterized by the hiring notification must. In particular, it consists of employee's personal data, the date of hire, the end date if the employment is not permanent, the type of contract, the professional qualification and the financial, the regulatory conditions applied and the ATECO sector to which the firm belongs to. The previous richness of information allowed the authors to analyse and compare the behaviour and results of companies belonging to different sectors.

To conclude, the following study offers an analysis of the labour demand in Italy's Emilia-Romagna region, with particular reference to the skills associated with the individual professions over the last decade and obtained through the use of data mining techniques and ESCO as secondary source. The analyzed period started with the end of the economic and financial crisis that began in 2008 and concluded shortly before the crisis resulting from the SARS-CoV-2 COVID-19 pandemic. A period, therefore, in which it is easier to identify medium-long-term trends linked to technical progress and not associated with other phenomena of a cyclical and/or structural nature.

The paper is organised as follows. Paragraph 4.2 discusses the Smart Specialisation Strategy in Emilia-Romagna, highlighting its logic and production-chain based structure. Paragraph 4.3 describes the data sources and sets out the methodological aspects of the analysis. Paragraph 4.4 presents an analysis of the aggregate labour demand for the whole region and compared between the different sectors. Paragraph 4.5 examines, in a comparison by sector, the development of professions and skills, highlighting those which have been most positive and most negative over the

[25] https://excelsior.unioncamere.net/

course of the decade. Paragraph 4.6 offers an analysis of the impact of Industry 4.0 technologies on professions and skills, highlighting the most relevant for each sector. Paragraph 4.7 provides a conclusion.

4.2 The Smart Specialisation Strategy and Clust-ER

As is well known, the European Commission required the adoption of the intelligent specialisation analytical system (Smart Specialisation) and the development of a strategy for its realisation as a condition for the development of cohesion policies of the Regions and Member States, to be financed with the Structural and Investment Funds for the period 2014-2020. The concept of smart specialisation is useful in order to provide consistency to certain requirements for the effectiveness of structural policies through their focus, including taking into account the results of previous programming and their critical issues through policy learning. Specifically, the Cohesion Policy 2014-2020 programming cycle provides, as an ex-ante condition for the use of EC resources, for national and/or regional authorities to develop research and innovation strategies for 'smart specialisation'[26], in order to allow more efficient use of Structural funds and an increase in synergies between EC, national and regional policies. The regions of all Member States are called upon to draw up a document to outline their strategy, starting out from their resources and capabilities, identifying the competitive advantages and technological specialisations most consistent with their innovation potential and specifying the public and private investments needed to support this strategy.

Following the preliminary work of the S3 Forums, Clust-ERs[27] made their debut in 2018, identifying themselves as one of the main elements of the Smart Strategy. The Clust-ERs are associations – one for each area of the regional S3 – of public and private bodies: research laboratories, companies, training institutions, innovation centres of the High Technology Network. Each Clust-ER is a community structured to share ideas, skills, tools and resources to support the competitiveness of S3 areas of the regional production system: agrifood, building and construction, mechatronics and motor engineering (which make up Priority A of the S3), culture and creativity[28], health and well-being (Priority B), energy and sustainability

[26] Main sources: https://fesr.regione.emilia-romagna.it/s3 and http://www.regione.emilia-romagna.it/s3-monitoraggio/about.html

[27] Main sources: https://fesr.regione.emilia-romagna.it/notizie/2018/giugno/clust-er-cosa-sono-e-come-aderire e https://www.retealtatecnologia.it/clust-er

[28] The scope of this sector may be defined primarily by the following activities: publishing industries; cinematographic and musical productions (audio-visual); activities related to the management, conservation, restoration and use of cultural heritage; creative and interactive digital industries; production of games and musical instruments; amusement parks; entertainment,

(Priority C) and innovation in services (Priority D). The Clust-ER project received support from POR FESR 2014-2020. Within these communities, laboratories and businesses work together according to the open innovation model to identify opportunities for partnerships and exploitation of research results. The Association promotes the development of shared projects and participation of its members in national and international calls and funding programmes. The region asked the Clust-ERs in particular to identify the priority goals for each area of specialisation of the S3 and with reference to the sectors/supply chains of greatest value for the regional economy (in terms of turnover, employment and competitive positioning in the international context). These objectives allow the definition of, among other things, the priorities on which to focus regional interventions for the last three years of the 2014-2020 schedule, and the Clust-ERs are configured as the brain of the production chains behind the regional production system.

Within this system, solid information on individual production chains and their evolution, also in terms of labour demand trends, takes on primary importance in the design of regional policies.

4.3 The data

4.3.1 Mandatory notifications

The collaboration of ART-ER, in-house company of the Emilia-Romagna Region, and the Emilia-Romagna Labour Agency, made it possible to build a suitable dataset for the analysis of the evolution of labour demand for the period 2008-2017. The dataset contains data on mandatory notifications sent to regional employment centres. Mandatory notifications on employment relationships are communications from employers (both public and private) which contain detailed information on each employment relationship and must be transmitted for all new hires and on transformation, extension and termination of the employment relationship. In addition to subordinated employment relationships, self-employment relationships of a coordinated and continuous form must be notified, including project-based work, cooperative workers, associates with job-based contributions, traineeships and any other type of similar work experience. The hiring notification must, in particular, give the employee's personal data, the date of hire, the end date if the employment is not permanent, the type of contract, the professional qualification and the financial and regulatory conditions applied.
In the analysis presented here, the employment relationships examined were drawn

show-business and cultural activities; design and communication services; creative services related to fashion and furniture; tourism linked to culture, entertainment and entertainment. (https://fesr.regione.emilia-romagna.it/s3) up to obtain a record for each employment relationship (always with the date of recruitment, showing the numbers of any transformations or extensions and, in the case of a completed relationship, the date of termination). The extension of the dataset in terms of economic activity (of the undertaking in which the employment relationship is active) shall include: manufacturing activities (ATECO2007 codes from 10.11.00 to 33.20.09), supply of electricity, gas, steam and air conditioning (ATECO2007 codes from 35.11.00 to 35.30.00), water supply; sewage systems, waste management and sanitation activities (ATECO2007 codes from 36.00.00 to 39.00.09), construction (codes ATECO2007 from 41.10.00 to 43.99.09), part of the information and communication services (ATECO2007 codes from 62.01.00 to 63.99.00), part of the professional, scientific and technical activities (ATECO2007 codes from 70.10.00 to 74.90.99) and part of the activities of renting, business support agencies and travel agencies (ATECO2007 codes from 78.10.00 to 82.99.99). The total number of employment relationships thus obtained was 3,123,108, of which 2,547,762 are terminated, while 575,346 were still active at the end of the analysed period (December 2017).

The economic activity codes of the companies present in the dataset allow the identification of the sectors in which the business participates (by associating the company's ATECO 2007 code with the sectors that include that code). In this way it is possible to analyse and compare the behaviour and results of companies belonging to the individual sectors. The criteria adopted for the identification of the sectors are those defined by the region in the Smart Specialisation Strategy. All known limits apply in the ability to account for the relationship between companies that cross the boundaries of the individual sectors.

4.3.2 Professions and skills in international statistics

The definition of what a skill truly is, represents a huge debate and it is deeply affected by the polysemy that characterize the term itself (Benadusi et al., 2018). Some authors suggest that the first distinction to be outline is the one between skill and performance (Chomsky, 1965), where the former indicates an individual potential, innate and universal. For other authors, skills could not be merely considered personal attributes, but are a mix of actions and relationships intertwined with organizational and social dimensions (Benadusi et al., 2018). For this reason, professional knowledge could be created, transmitted, interpreted. Therefore, skills are generated and developed but can also be impoverished, degraded and renewed in the circuit of educational, training and work experiences, throughout the life and in the passage between different contexts, following the socioconstructivist perspective (Jonnaert, 2009; Benadusi et al., 2018). Although polysemy is undoubted, skills can be categorized as resources to act, as devices that allow actions and are formed in action (Benadusi et al., 2018). Moreover, skill can be learned by immersing in practice, and during the processes of family, educational and professional socialization. (Dubar, 1991). For those reasons, they could be gain through the process of formation of individual and social identities (Ajello 2002; Pontecorvo et al., 1995). For what concerns the present work, when citing the term skill, the authors are aligned to ESCO and EQF definition (i.e. "skill means the ability to apply knowledge and use know-how to complete tasks and solve problems.") where the term skill refers typically to the use of methods or instruments in relation to defined tasks (European Commission, 2020).

As regards skills analysis, the following scheme shows which information has been selected and which additions have been made in order to obtain the final result. Every observation of the SILER database, described above, is representative of an employment movement associated with the individual's 5-digit ISTAT professional classification, adopted from 2011 and aligned with the European ISCO-08 standard. Specifically, the ISTAT professions variable consists of 5 digits, each corresponding to a grouping level (Figure 23):

Figure 23: Professions classification tree structures

The legend for the ISTAT Main Groups is given below:

1. Legislators, entrepreneurs and senior management

2. Intellectual, scientific and highly specialised professions

3. Technical professions

4. Executive professions in office work

5. Qualified professions in commercial activities and services
6. Craftsmen, skilled workers and farmers
7. Plant operators, fixed and mobile machinery workers and vehicle drivers
8. Unskilled professions

To trace the skills associated with each working relationship in the SILER [29] archive, a crosswalk was carried out with the ESCO (European Skills, Competences, Qualifications and Occupations) database which incorporates the European standard for skills, qualifications and occupations. ESCO has a structure similar to that of a dictionary, describing, identifying and classifying relevant professions, skills and qualifications and EU education and training. Every observation of the ESCO database is identified by an ISCO-08 code. The connection table[30] between ISTAT 2011 and ISCO-08 can similarly combine the SILER database and the ESCO database. The process is briefly outlined in Figure 24.

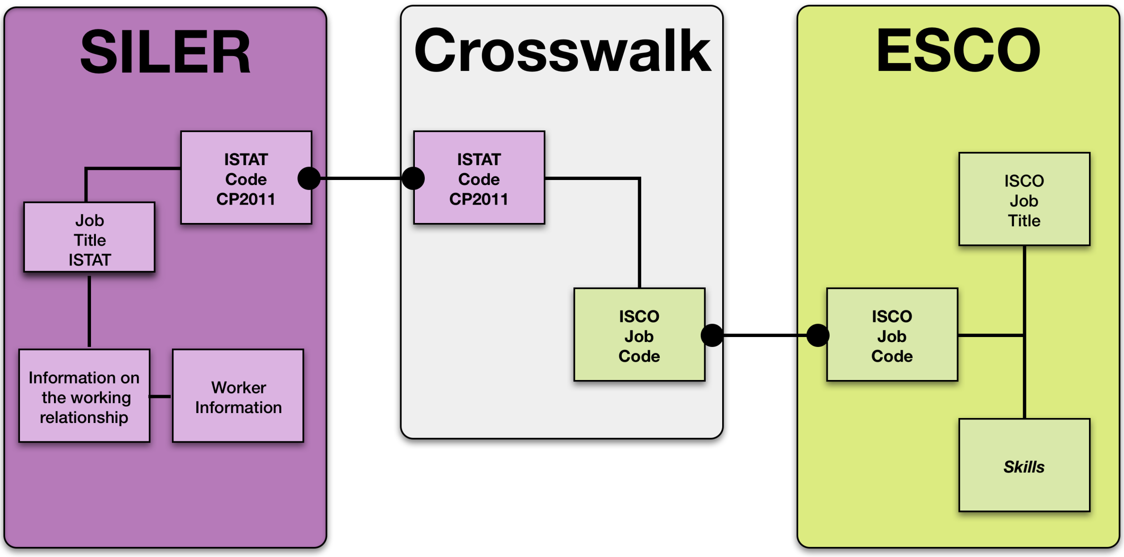


Figure 24: Dataset scheme and variables used for the calculation of the employment balance

Through the crosswalk it was possible to express the skills characterising the different professions. The employment trend was then first observed, and then the skills characterising the occupational profiles with a markedly positive or markedly

[29] In the dataset created, there is no employment relationship featuring a profession of Main Group 9 – Armed Forces.
[30] https://www.istat.it/it/files/2011/04/raccordo$_Isco08_CP2011.xls$

negative employment balance (by individual sector and in terms of comparison). Finally, a search was carried out for digital skills only among the profiles with the most positive employment balance, which also identified how they are distributed across the different sectors. This was made possible through the help of a dictionary enriched with Industry 4.0 technologies (Chiarello et al. 2018), which will be described in detail in the following sections.

4.4 Sectors in the regional economy

4.4.1 The quantitative relevance of sectors (2008-2017)

Before focusing on the comparison between the sectors, it is appropriate to focus on the relative weight of the different sectors in terms of workers and enterprises in the regional economy during the period under consideration (2008-2017).
The official data from the institute of statistics (ISTAT,Labour Force Survey[31]) tell us that in Emilia-Romagna in 2008 there were about 1,950,000 people in employment. The data at our disposal from the SILER source, for the same year, provide us with information for about 1,650,000 employed persons. For the other years analysed in this survey, the report is broadly the same: information is therefore available on about 85% of those employed in the region. Of these workers, approximately 1,200,000 (in 2008) were employed in enterprises which are part of the sectors of the Clust-ERs (72% of the sample and 62% of the total employed in the region). The three most representative sectors for the region's economy, namely Agrifood, Building and Construction, and Mechatronics and Motor, accounted for 23.2%, 22.1% and 19.4% of the total employed, respectively, in 2008. The other four sectors together accounted for 28.6% of workers. Table 18 shows the number of workers for three selected years (beginning, middle and end of the period).

	Workers (2008)	Workers (2012)	Workers (2017)	% on the total number of workers in the supply chain (2008)	% on the total number of workers in the supply chain (2012)	% on the total number of workers in the supply chain (2017)
Agrifood	383.988	373.793	431.044	24,91%	26,38%	28,31%
Build	365.077	309.693	297.757	23,68%	21,85%	19,56%
Mech	319.588	277.803	304.510	20,73%	19,60%	20,00%
Health	134.307	132.668	150.710	8,71%	9,36%	9,90%
Create	185.995	169.507	171.879	12,06%	11,96%	11,29%
Innovate	64.090	71.002	81.758	4,16%	5,01%	5,37%
Greentech	88.707	82.711	84.900	5,75%	5,84%	5,58%
Total	1.541.752	1.417.177	1.522.558	100,00%	100,00%	100,00%

Table 18: Number of employees in 2008, 2012 and 2017

Regarding the interpretation of Table 18, it's important to outline that the total number of workers obtained from the sum of the workers in the individual sectors is greater than the actual number of workers present in our sample, since some companies and, therefore, their workers, contribute to more than one sector. For example,

[31] https://www.istat.it/it/archivio/8263

Year	Permanent Contract	Temporary Contract	Administration Contract	Apprenticeship	Parasubordinate	Total
2008	835.322,50	172.496,50	23.101,59	46.047,01	38.384,42	1.115.352,02
2009	821.729,30	162.454,20	16.258,38	43.044,28	36.246,16	1.079.732,32
2010	804.618,80	162.948,60	19.525,36	39.625,07	36.350,73	1.063.068,56
2011	797.117,80	167.655,30	24.960,63	37.979,87	37.760,54	1.065.474,14
2012	795.496,50	160.111,70	25.411,42	36.919,95	37.327,93	1.055.267,50
2013	786.903,40	155.973,40	28.024,36	36.605,09	31.049,70	1.038.555,95
2014	773.703,30	160.695,40	31.545,49	37.670,08	30.016,81	1.033.631,08
2015	775.486,70	160.157,40	35.826,63	38.989,44	26.926,55	1.037.386,72
2016	798.285,20	162.693,70	40.078,14	41.676,57	17.442,14	1.060.175,75
2017	777.011,90	175.339,50	47.959,79	47.875,21	16.987,45	1.065.173,85
2017/2008 (%)	-6,98%	1,65%	107,60%	3,97%	-55,74%	-4,50%

Table 19: ULA trend over time by type of contract

the manufacture of ceramic sanitary ware (ATECO 23.42.00) is part of both the building and construction sector and the health and well-being sector.

Some observations can be made from Table 18 on the development of employment in the individual sectors. In particular, the Agrifood (Agrifood), Health (Health and Well-being) and Innovate (Innovation in Services) sectors recorded a significant increase in the share of workers from the total over the period (from 37.77% to 43.58% between 2008 and 2017, when aggregated). On the contrary, the sector which lost the most workers (as a share of the total, but also in absolute terms) is Building and Construction (Build). The other sectors analysed did not see significant change in their share of workers from the total over the period under analysis.

Taking into account the number of Annual Working Units (ULA[32]), it can be noted that the pre-crisis employment level was not yet reached by the end of the period: compared with about 1,115,000 annual working units in 2008 (to be attributed to the previously mentioned 1,650,000 workers), in 2017 there were indeed still only 1,065,000 thousand ULAs (about 4.5% less). The reduction in the amount of work was not uniform in the different types of contract (Table 19 and Figure 25): "parasubordinate" (or contrived self-employment terms) and work under permanent contracts fell significantly as a percentage, while work under temporary contracts and apprenticeship and placement work returned substantially to pre-crisis levels while annual work units under staff-leasing contracts more than doubled.

[32]Unit of measurement representing the amount of full-time annual work: it is, in other words, a virtual estimate of the number of people employed if it were assumed that everyone was working on a full-time basis.

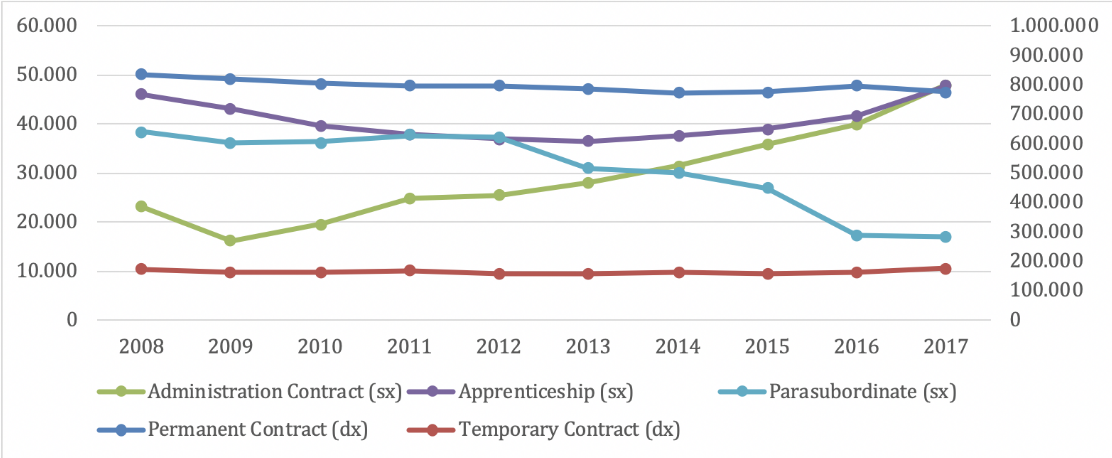


Figure 25: ULA trend over time by type of contract. The ULA number of the left vertical axis refers to the staff-leasing, apprenticeship and placement, and "para-subordinate" types. The number of ULAs of the right vertical axis refers to the permanent and temporary contract types.

With regard to the ULAs used by the companies in the analysed sectors, and observing their evolution over time (Table 20), we can note that the trends which have emerged with regard to the number of workers were largely confirmed (see Table 18).
The following Tables 20 and Figure 26 show the trend of the Annual Work Units by sector, for the analysed period of time.

Year	Agrifood	Build	Mech	Health	Create	Innovate	Greentech	Total
2008	240.695	258.622	243.340	97.631	111.446	43.423	68.120	1.063.275
2009	239.303	246.126	227.790	98.823	108.643	44.471	66.288	1.031.444
2010	238.723	238.100	220.793	99.877	107.503	45.142	65.272	1.015.409
2011	242.280	235.671	223.589	100.346	108.240	47.146	67.208	1.024.480
2012	243.952	228.556	222.715	99.906	108.340	48.534	66.270	1.018.272
2013	244.284	220.009	219.326	99.220	106.072	48.443	64.685	1.002.038
2014	248.361	212.038	219.364	99.567	105.022	50.298	63.720	998.370
2015	251.771	208.136	221.540	100.574	105.023	52.324	63.545	1.002.913
2016	260.972	210.224	228.543	103.442	107.528	52.821	64.544	1.028.075
2017	266.283	207.979	232.474	105.921	107.732	53.676	65.737	1.039.801
2017/2008 (%)	10,63%	-19,58%	-4,47%	8,49%	-3,33%	23,61%	-3,50%	-2,21%

Table 20: ULA trend over time by sector - Absolute values given without decimal places for clarity. The total number of workers obtained from the sum of the workers in the individual sectors is greater than the actual number of workers present in our sample, since some companies and, therefore, their workers, contribute to more than one sector. For example, the manufacture of ceramic sanitary ware (ATECO 23.42.00) is part of the building and construction sector and the health and well-being sector.

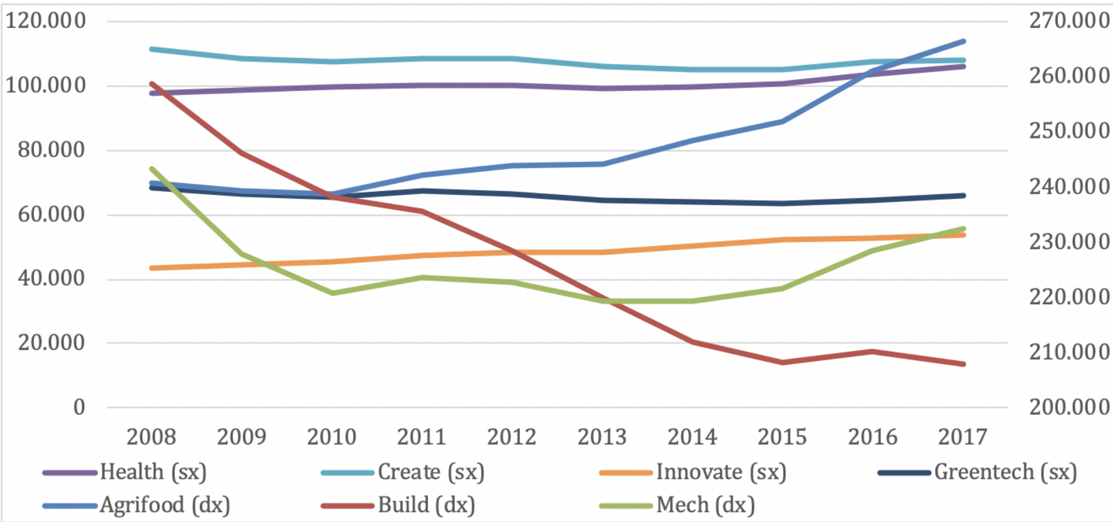


Figure 26: ULA trend over time by sector. The ULA number of the left vertical axis refers to the Health, Create, Innovate and Greentech sectors. The ULA number of the right vertical axis refers to the Agrifood, Build and Mech sectors.

Figure 26 helps to identify the most significant trends. The Building and Construction sector lost almost 20% of the ULAs employed at the beginning of the period; on the contrary, the Agrifood sector saw an increase of more than 10% in the ULAs used; finally, in the Innovation in services sector, the increase exceeded

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Number of firms	75.473	73.782	73.568	73.726	73.994	73.041	72.033	73.016	72.847	71.819
Number of firms part of a sector	56.005	54.564	54.266	54.396	54.627	53.900	53.134	53.841	53.656	52.838
Number of workers part of a sector	1.207.951	1.110.638	1.104.413	1.112.147	1.098.412	1.073.271	1.082.485	1.103.133	1.127.867	1.173.278
Average number of workers for firms part of a sector	21,569	20,355	20,352	20,445	20,107	19,912	20,373	20,489	21,020	22,205

Table 21: Number of enterprises and workers

Sector	Turnover
Agrifood	0.529
Build	0.408
Mech	0.333
Health	0.391
Create	0.458
Innovate	0.431
Greentech	0.363
Total	0.446

Table 22: Turnover by sector – Average 2008-2017

23 percentage points. The dynamics of other regional sectors are less clear. The Mechatronics and motor sector, for example, behind a reduction of about -4% of the ULAs used, hides different trends in different sub-periods: a sharp decline in the first years of the range analysed, and a recovery from 2014 onwards. Trends in the automotive sector largely explain this.

The number of enterprises operating in the regional territory which are part of at least one sector is given in Table 21.

These data allow the turnover to be calculated for each sector. The definition of turnover used here is linked to the Annual Labour Units (ULA) and is equal to the number of ULAs handled during the year by the company (incoming and outgoing) divided by the total number of ULAs. The results of this are given in Table 22.

Turnover has values of between zero and one. In the first case (turnover=0) there was no movement of workers (ULA) within the enterprise; on the contrary, a turnover value of 1 means a number of incoming and outgoing ULAs equal to the total number of ULAs in the period. The first important figure is the high average value of the exchange in all sectors: in all sectors, on average, one in three workers is replaced over the course of the year. Increasingly in the region (although this

Sector	Average duration of contracts terminated in the period 2008-2017 (in days)	Average duration of contracts terminated in the period 2008-2017 (compared to the average)
Agrifood	615,27	0,81
Build	808,66	1,06
Mech	958,62	1,26
Health	861,55	1,13
Create	794,48	1,04
Innovate	491,49	0,64
Greentech	902,25	1,18
Total	762,75	1,00

Table 23: Average duration of contracts terminated in the period 2008-2017

applies in all major developed economies) work changes back from an "almost fixed factor" to a "variable factor", at least in terms of some of its components, depending on the evolution of product demand. We will return to this topic in greater depth in the following pages, analysing the twin variable of the turnover rate, that is the average duration of the employment relationships. For now, it is sufficient for us to emphasise that the sectors with a higher turnover are those of Agrifood, in which seasonal work exerts significant weight, and the Culture and Creativity (Create) sector. The sectors with a relatively low turnover rate are Mechatronics and Motor (Mech) and Energy and Sustainability (Greentech).

The average duration of contracts terminated in the period 2008-2017, by sector, is shown in Table 23 and Figure 27.

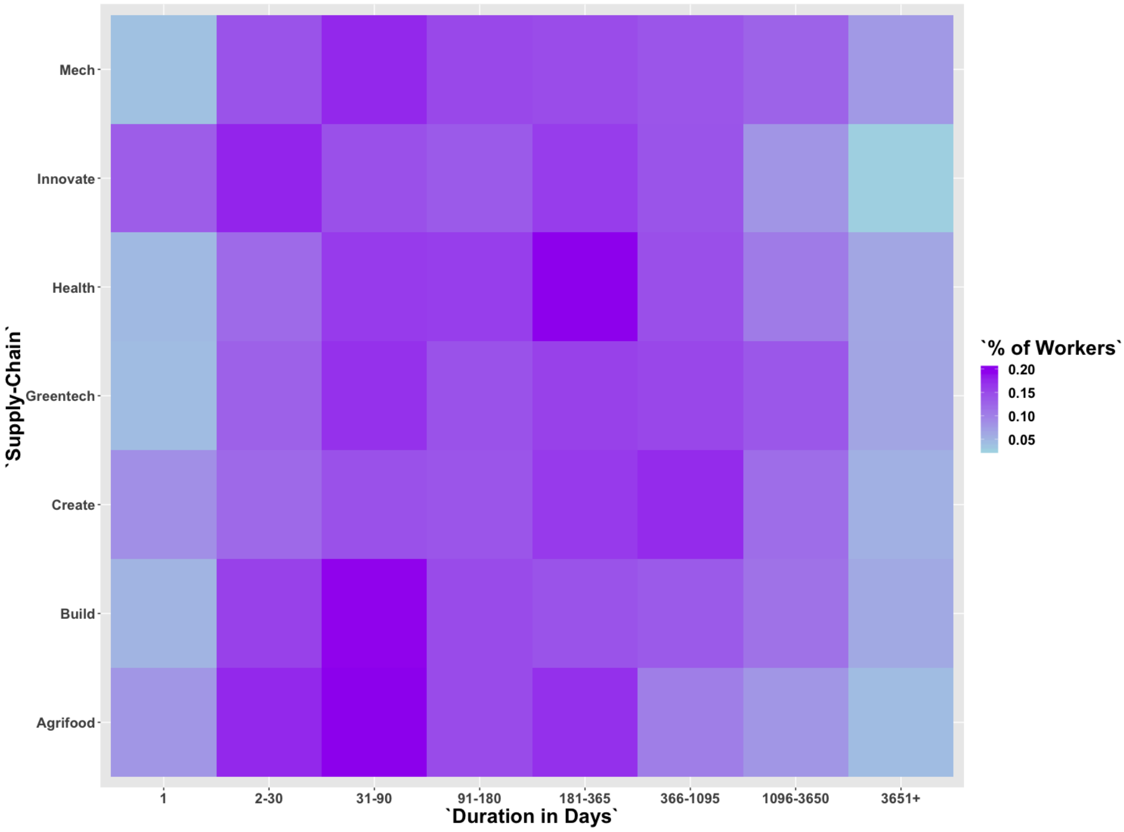


Figure 27: Concentration of the percentage of workers by sector and duration of contract

The analysis of the duration, as was easy to expect, confirms and qualifies what emerged from the analysis of turnover. The relatively long duration is noted for the mechatronics and motor sector (Mech), Energy and Sustainability (Green) and Health and Well-being (Health). On the contrary, the shortest contracts were found in the Agrifood (Agrifood) and, above all, Innovation in Services sectors. The average duration of contracts in the Mechatronics and Motor sector which ended in the period 2008-2017 was almost double the Innovate sector (nearly two years eight months versus just over one year and four months, respectively). Significantly, relatively longer averages were recorded in the Mech and Build sectors, which, as noted, experienced a sharp reduction in employment at the beginning of the period: the relatively longer duration could be evidence of mass dismissal of workers of significant seniority. The short durations in the Agrifood and Innovate sectors are probably ascribable to very different factors. In Agrifood, as previously noted, part of this trend should be attributed to seasonal work. Both sectors, however, are made up of activities undergoing growth which, perhaps more than the others, make use of short-term contracts. In some cases, for the higher-level professional profiles, they may also result from poaching (attempting to lure away workers with the highest qualifications).

Data on turnover and duration should be read with great care. Firstly, we need to be aware that the mandatory notifications and the dataset that derives from them, like all other labour archives from administrative sources, are regulated by the Italian civil code. Although there is substantial continuity, certain changes in the ownership or legal form of the company imply the closure of the old entity and the creation of a different one. In terms of employment contracts, this means that the old contract is extinguished and a new one created, even if the job has not changed, nothing has altered in the relationship between the worker and the company, and the employee does not perceive the transformation. The contract is terminated and recreated for purely administrative reasons. This has a significant effect on both the turnover and duration measured. It tends, indeed, to increase the turnover rates recorded in individual jobs (those in which the company has undergone a legal transformation as outlined above). At the same time, where there is transformation, the prior duration of the contract is reset, with the effect of shortening the durations. Since such phenomena are recurring over the life of the company, long durations tend to be significantly underrepresented.

Secondly, high measured rates of labour rotation do not mean that there are few employees with a high level of seniority. On the contrary, they certainly continue to be the majority of those employed in the private sector as well. It simply means that most of the movements — those that the dataset allows us to measure accurately – are generated by workers with short-term contracts. As is well known, high corporate seniority and high turnover are two complementary measures in the analysis of labour market flows and, under certain conditions, can coexist. In particular, in dual labour markets, as Italian labour markets typically are, turnover can be generated by a small but very mobile component compared to a much larger proportion of workers with stable and long-term employment relationships. Both components can enjoy very differentiated wage and working conditions[33]. In the background are all the nuances which have characterised the debate on flexibility/precariousness of Italy's labour markets over the last few decades. This is particularly true as regards young people.

These warnings are necessary for reading data that can be provided by an archive

[33]Among the many studies concerning the flow analysis of labour markets, with particular reference methodological aspects and to Italy, we recall: Contini (2002 and 2005); Contini and Revelli (1997); Contini and Trivellato (2005); Haltiwanger, Manser and Topel (1999), Solinas (1990); Torelli and Trivellato (1993); Trivellato (2011). Here, we will simply remind the reader that turnover is a relationship between a flow and a stock: the second counts the heads, the first movement; and that an individual can generate multiple movements during the observation period.

built on mandatory notifications. It can perhaps be said that what the archive reveals most accurately is what changes, the movements "on the margins" of the labour markets. The results presented below will be analysed from this point of view. On the other hand, it provides a much less focused view of stable/continuous relationships: those who remain in their workplace without generating movements.

4.4.2 A comparative analysis of the sectors in Emilia-Romagna

After a summary of the numbers and turnover, the workflows for the individual sectors during the period under consideration are analysed below, using mandatory reporting data.

Figure 28 shows the change in employment between 2008 and 2017 by individual sector. All sectors, with the exception of Health, suffered the effects of the 2008 crisis, reaching their nadir in 2009 (2011 for Health and well-being). Although there was a recovery for almost all the sectors in 2010-11, the trends fluctuated, with employment balances deteriorating in 2012-13, improving from 2014-16 and then dropping again after 2016. Only Greentech recorded an increase in employment for 2017 compared to 2016.

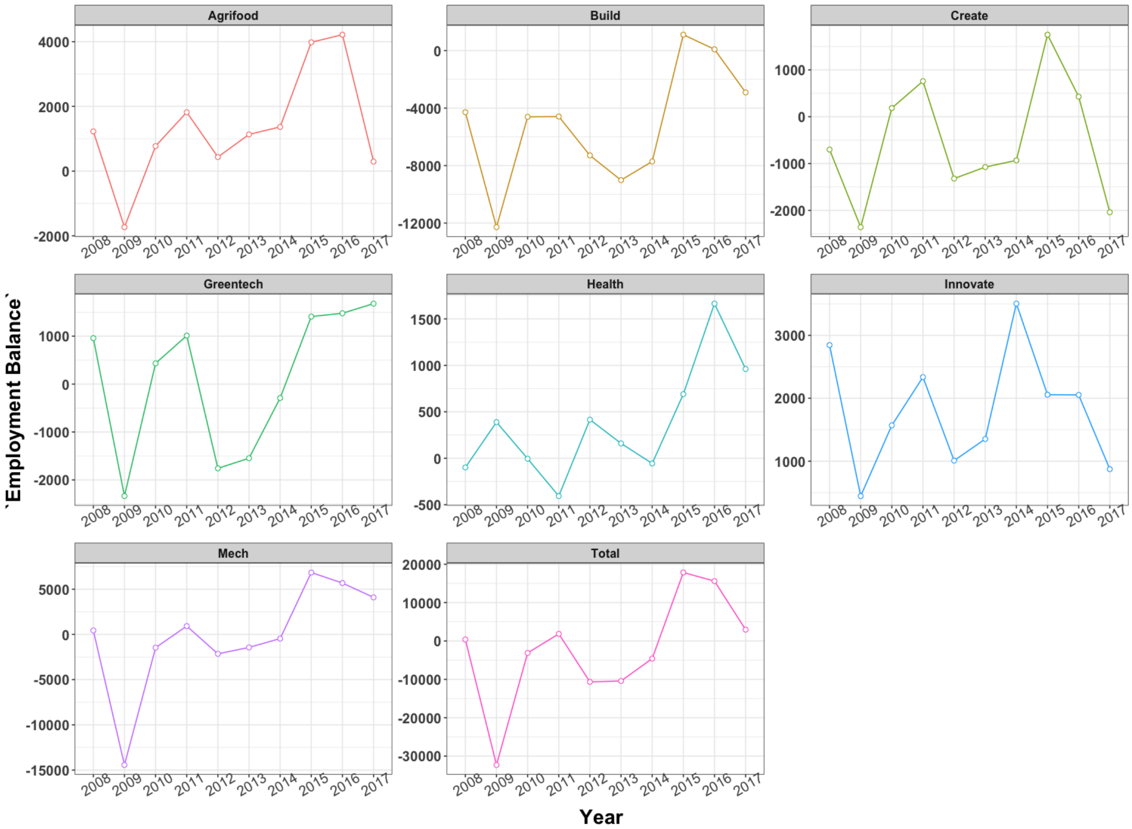


Figure 28: Trends in the employment balance in individual sectors from 2008 to 2017

Table 24 and Figure 29 provide a more in-depth view of the trends. The Construction sector featured the worst employment balances, which were always negative except for the two-year period 2015-16. In the ten years taken into consideration, workers in this sector fell by more than 50,000. Mechanics, Mechatronics and motor suffered a strong negative impact in 2009, followed by a period of settlement characterised by minor balances (2010-2014) and then a significant recovery after 2015 (which allowed it to limit losses in total employment to just under 2,000). The Health and Well-being sector shows a very peculiar dynamic compared to the others, as already observed, with a significant negative balance only in 2011 and without suffering particular repercussions, which could have been expected, from the effects of the earthquake of May 2012, which the biomedical district of Mirandola, among others, was heavily affected by. The Greentech, Create and Agrifood sectors did not change significantly over the period, with alternating positive and negative balances, although with a general improvement after 2014 and a worsening in 2017. The Innovation sector in services is worth special mention: it is the only one that has never experienced negative balances, and at the end of the period the total balance was

Sector	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
Agrifood	1.228	-1.731	774	1.820	435	1.134	1.366	3.981	4.215	295	13.517
Build	-4.284	-12.285	-4.609	-4.586	-7.294	-9.014	-7.715	1.106	90	-2.909	-51.500
Mech	437	-14.436	-1.464	920	-2.149	-1.441	-463	6.842	5.677	4.093	-1.984
Health	-99	389	-5	-407	416	160	-56	690	1.665	961	3.714
Create	-701	-2.358	185	759	-1.321	-1.075	-932	1.751	429	-2.040	-5.303
Innovate	2.846	446	1.570	2.337	1.009	1.353	3.505	2.057	2.054	874	18.051
Greentech	959	-2.337	433	1.012	-1.759	-1.547	-290	1.410	1.478	1.680	1.039
Total	386	-32.312	-3.116	1.855	-10.663	-10.430	-4.585	17.837	15.608	2.954	-22.466

Table 24: Net employment balance by year and sector

18,000 units greater than at the beginning of the period.

Figure 29: Total trend of the employment balances of the various sectors from 2008 to 2017

All sectors slowed down more or less markedly at the end of the period, highlighting that the regional production system, particularly manufacturing, was in difficulty even before the pandemic-induced crisis. In addition to specific sectoral problems and the effects of substitution with products from newly industrialised countries, German manufacturing performance and the outbreak of the US-China conflict, with its negative impact on world trade, are easy to blame.

As we have previously noted, the SILER dataset provides important information on the composition of labour demand. Below, by way of example, a number of aspects are highlighted. The first, which is not obvious, regards the numbers of women in the different sectors.

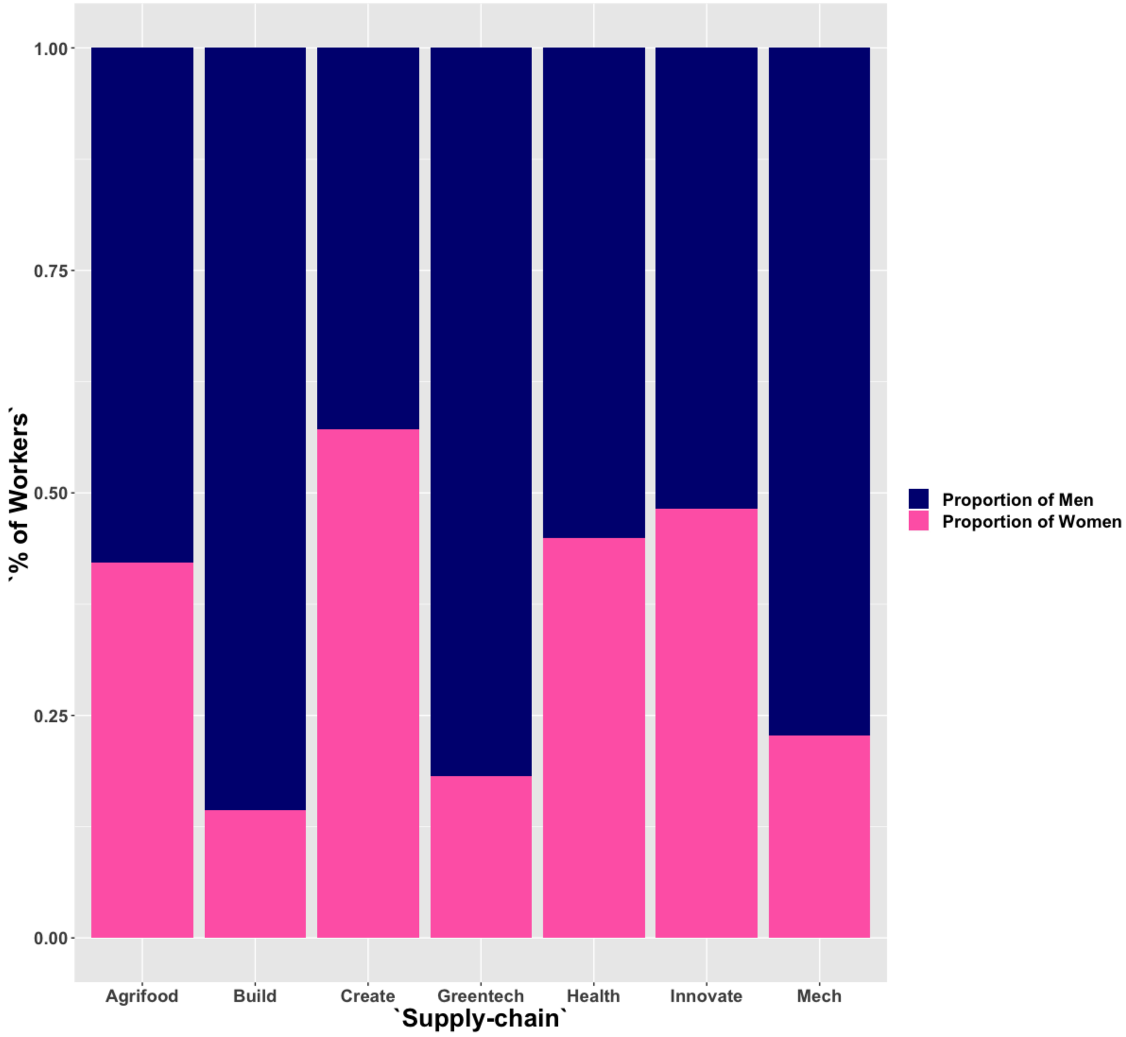


Figure 30: Percentage breakdown by employee gender in the period 2008-2017 by sector

Figure 30 shows the gender composition percentage of hires in the sectors. The Build, Greentech and Mech sectors were composed mainly of men; Create, on the contrary, had a prevalence of female workers. Health, Agrifood and Innovate showed a more balanced gender distribution. The comparison between Figure 30 and the previous Figure 29 allows an important aspect to be highlighted. Two of the sectors recording negative balances (Mech and, above all, Build) had a clear prevalence of men.

The training of workers has been an aspect of focus for observers and researchers of the labour markets for some decades now. An aspect of interest relates to the level of formal education. Figure 31 shows the concentration of the level of education achieved by single sector. Middle-school and high-school leaving certificates and no declared title appear to be the most common in all the sectors; the first, in particular, is highly concentrated in the Build and Greentech sectors. The Health sector, along with the Innovate, are those with the highest percentage of university degree holders.

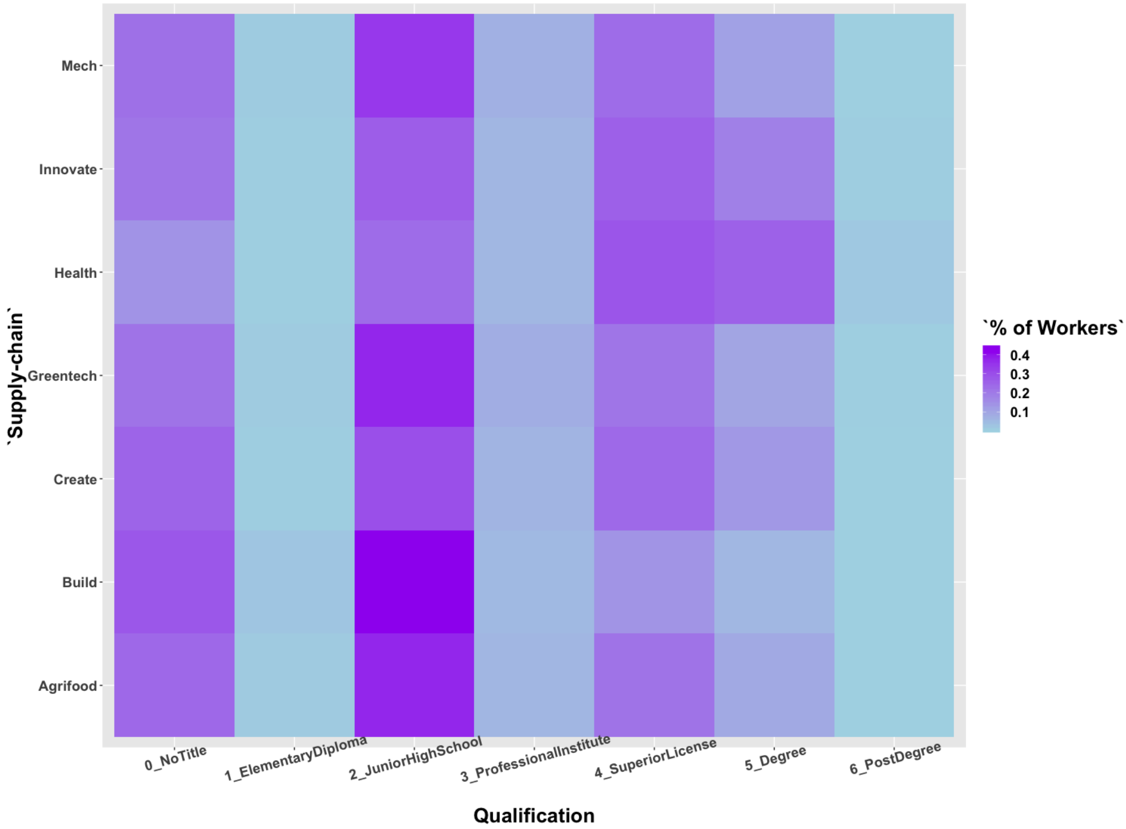


Figure 31: Concentration of the percentage of workers by sector and qualification

Again, the low level of school education is striking, even in sectors and supply chains requiring extensive use of advanced technologies. Particular attention should be paid to the development of recruitment and the employment balance of employees with high-school diplomas and university degrees during the decade under consideration. Figure 32 shows the employment balance for the regional production system as a whole.

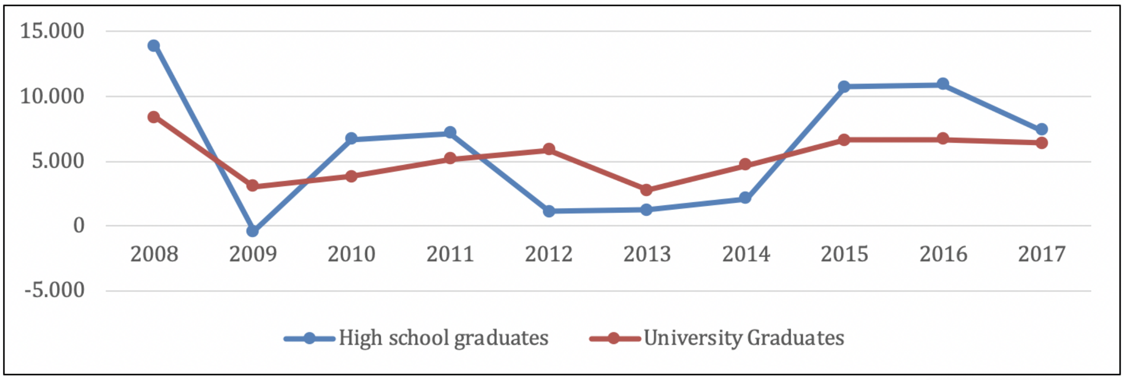


Figure 32: High-school diploma and university degree holders: employment balance for all regional production systems

High-school diploma holders (and certainly all workers with lower average levels of education) are far less protected than university degree holders when faced with contractions in demand. The adjustment of the labour market falls mainly on the former. This is also confirmed by examination of durations by qualification. Those with only middle-school education (the minimum mandatory level in Italy) tend to have monthly or quarterly contracts in almost all sectors, but in particular in Build, Greentech and Agrifood (Figure 32).

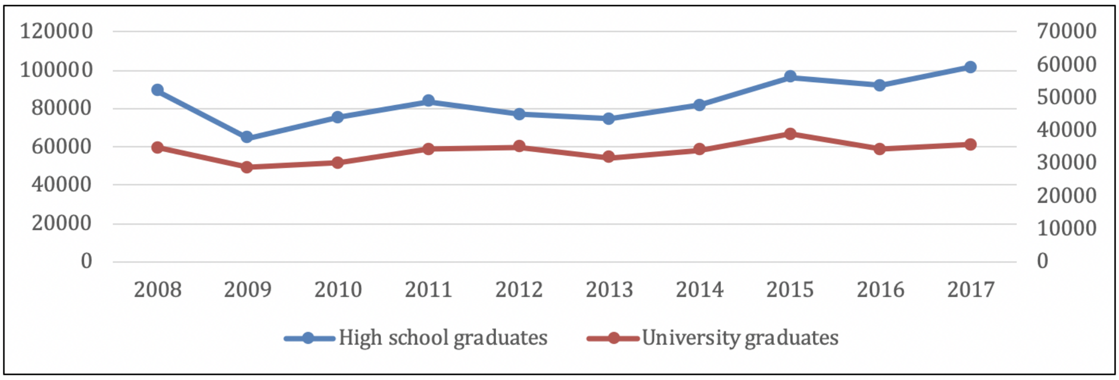


Figure 33: High-school diploma and university degree holders employed by all regional sectors

Perhaps the most important fact is that throughout the period, the recruitment of high-school diploma holders was consistently higher than that of university degree holders (Figure 34). Despite all the significant changes which are in progress, the regional economy, even at medium-high qualification levels, continues to rely more heavily on high-school graduates than on university graduates.

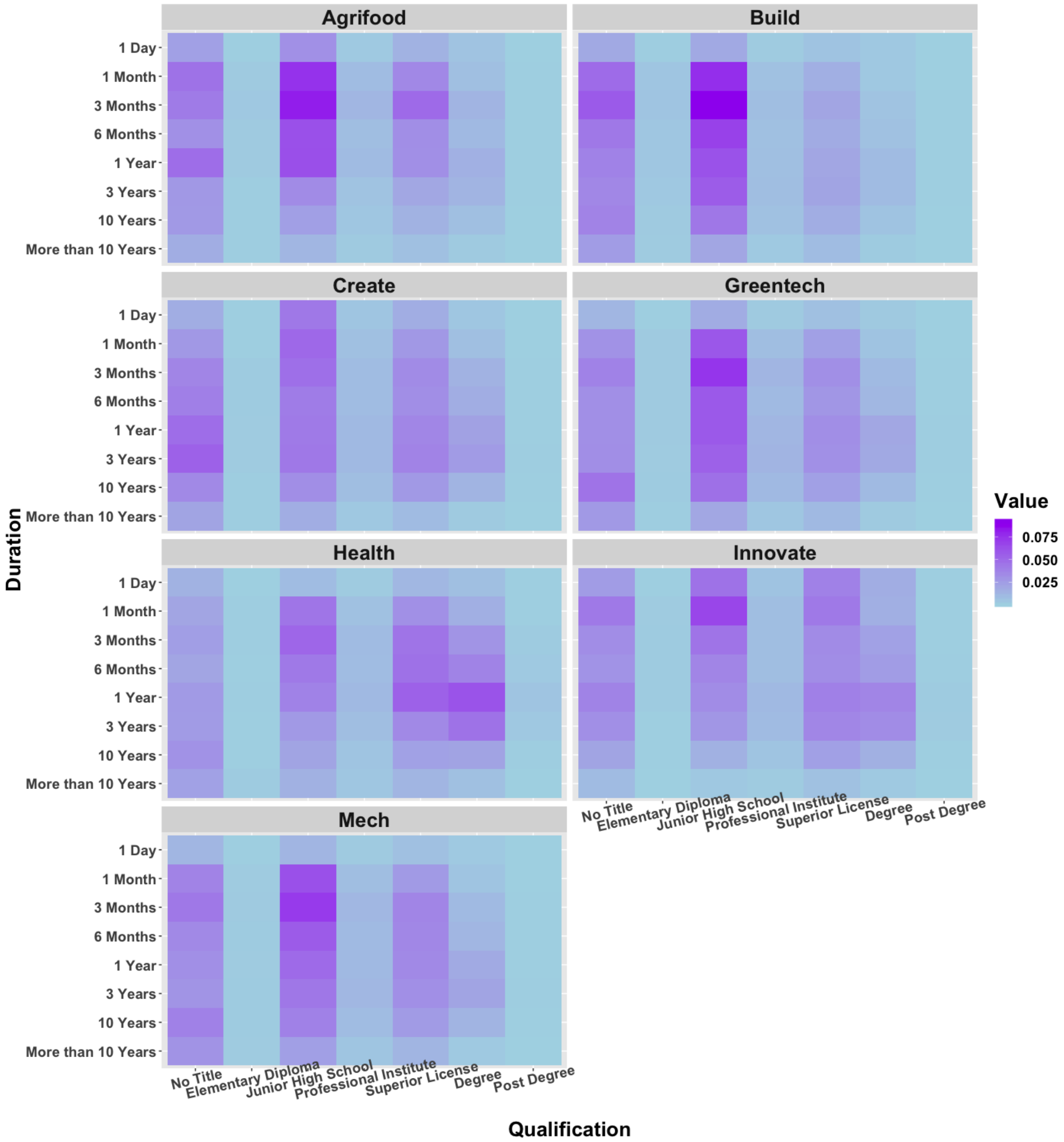


Figure 34: Relationship between employment contract duration and qualification by sector

A further aspect we feel it useful to highlight concerns the comparison between Italian and foreign workers. Figure 35 shows the number of hires and terminations by sector and citizenship (Italian or foreign). Almost one in four movements (incoming or outgoing) involves foreign workers: in particular, these workers generate a significant proportion of movements in the Building and Construction (almost one in three) and Culture and Creativity sectors. It is interesting that, for the whole period analysed, the negative employment balances (Build, Mech and Create, see Table 24) are due to a reduction in Italian rather than foreign workers. For example, in the Mechatronics and Motor sector, the overall balance of foreign workers is positive, albeit by little, compared to a decrease of more than 2,000 Italian workers in the relevant period. These data show that Emilia-Romagna's society, and even more so its labour market, is now multi-ethnic and open to the contribution of workers of many different backgrounds.

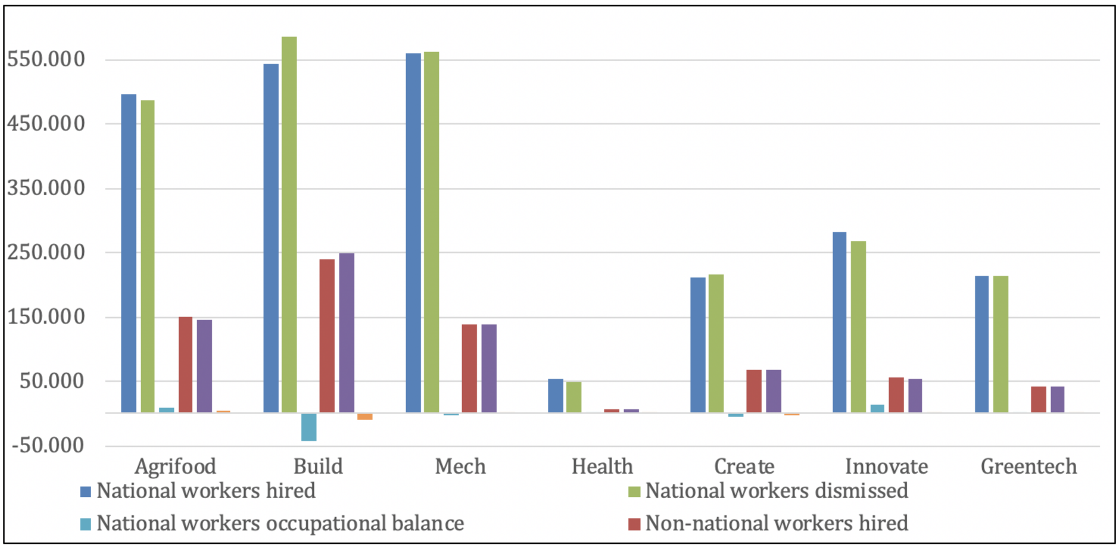


Figure 35: Incoming and outgoing movements and balance of Italian and foreign workers by sector – 2008-2017

One of the fundamental variables on which the topic covered in this paper reflects – and crucial to understanding the incoming and outgoing flows from individual jobs (and from individual sectors) – is occupation. Figure 36 shows the concentration of the type of profession (horizontal axis) for each individual sector.
Predictably, the Build, Greentech and Mech sectors have similar behaviours and very high concentrations of blue-collar workers, craftsmen and farmers; the number of qualified professions in commercial activities is negligible. Agrifood and Build also show a high concentration of unskilled professions. As regards Innovate, Health and Create, the most important professions are the technicians, intellectuals and office executives.

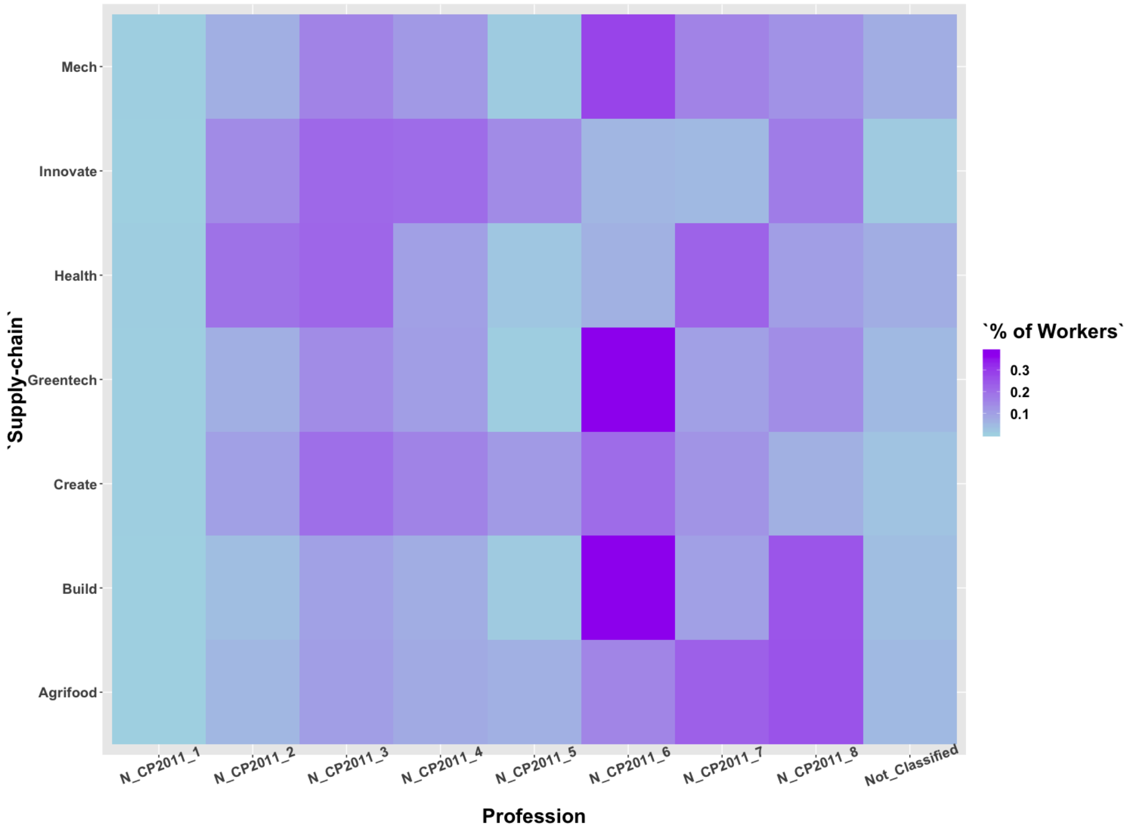
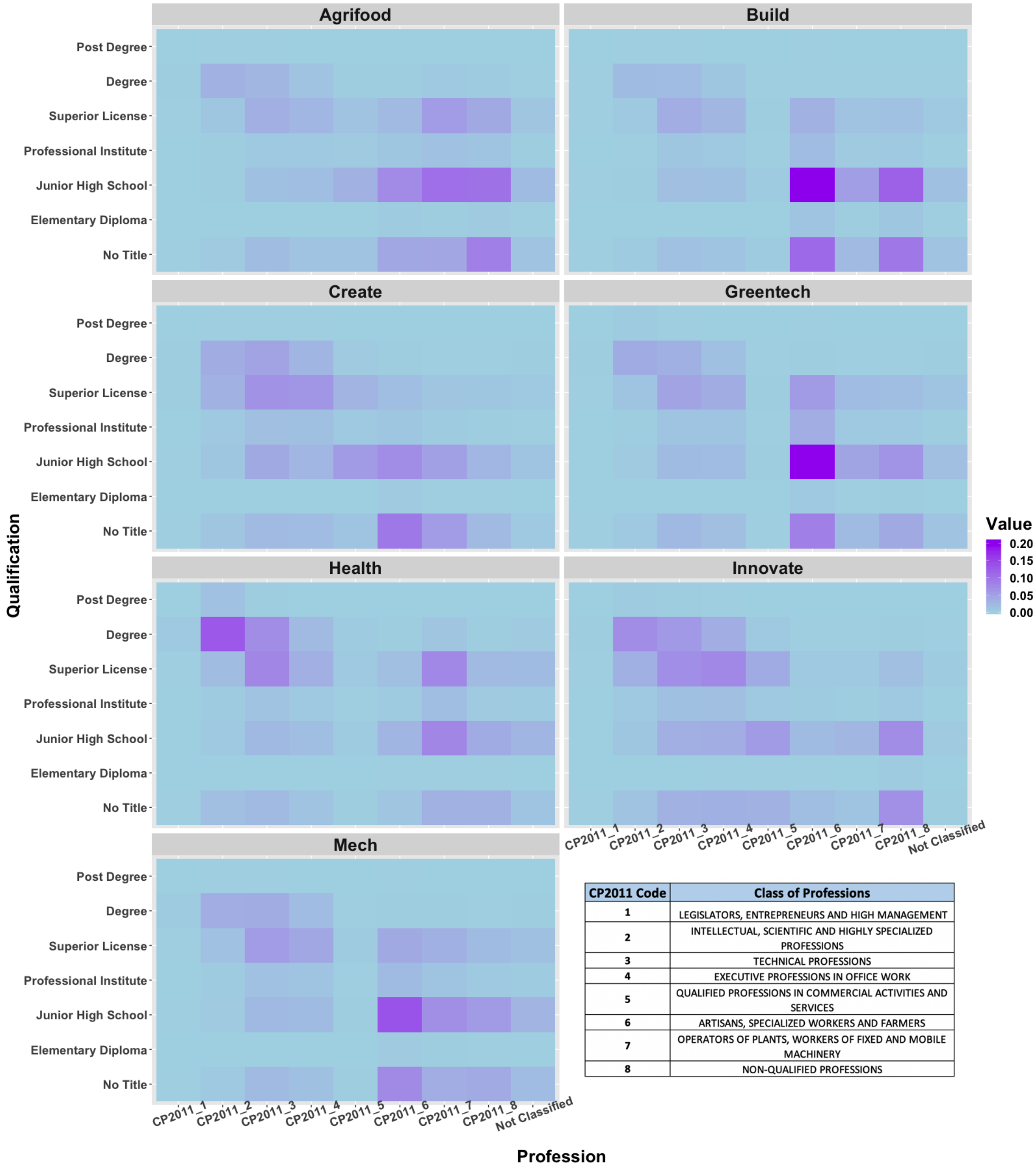


Figure 36: Concentration of workers by sector and ISTAT profession – Percentage values

37 shows the correlation between the ISTAT professional class and the highest educational qualification. Craftsmen, workers and farmers are particularly associated with a middle-school education in almost all sectors, but particularly in Greentech, Build and Mech. University degree holders correlate strongly with the intellectual and scientific professions, in particular in the Health and Well-being sector.



CP2011 Code	Class of Professions
1	LEGISLATORS, ENTREPRENEURS AND HIGH MANAGEMENT
2	INTELLECTUAL, SCIENTIFIC AND HIGHLY SPECIALIZED PROFESSIONS
3	TECHNICAL PROFESSIONS
4	EXECUTIVE PROFESSIONS IN OFFICE WORK
5	QUALIFIED PROFESSIONS IN COMMERCIAL ACTIVITIES AND SERVICES
6	ARTISANS, SPECIALIZED WORKERS AND FARMERS
7	OPERATORS OF PLANTS, WORKERS OF FIXED AND MOBILE MACHINERY
8	NON-QUALIFIED PROFESSIONS

Figure 37: Correlation between ISTAT professional class and academic qualification by sector

38 shows the correlation between the profession and the duration of the professional relationship in the individual sectors. In this case too, Build, Greentech and Mech exhibited similar behaviour, in which there is a strong correlation between blue-collar workers/craftsmen and quarterly (or in any case short) contract

duration. In the Innovation in Services sector, qualified professions have a high concentration of day-rate work. Similarly, in Agrifood, plant operators and professions have short-term contracts, in particular annual contracts. Both in Health and Innovate, high-skilled intellectual professions have a prevalence of annual or three-year contracts.

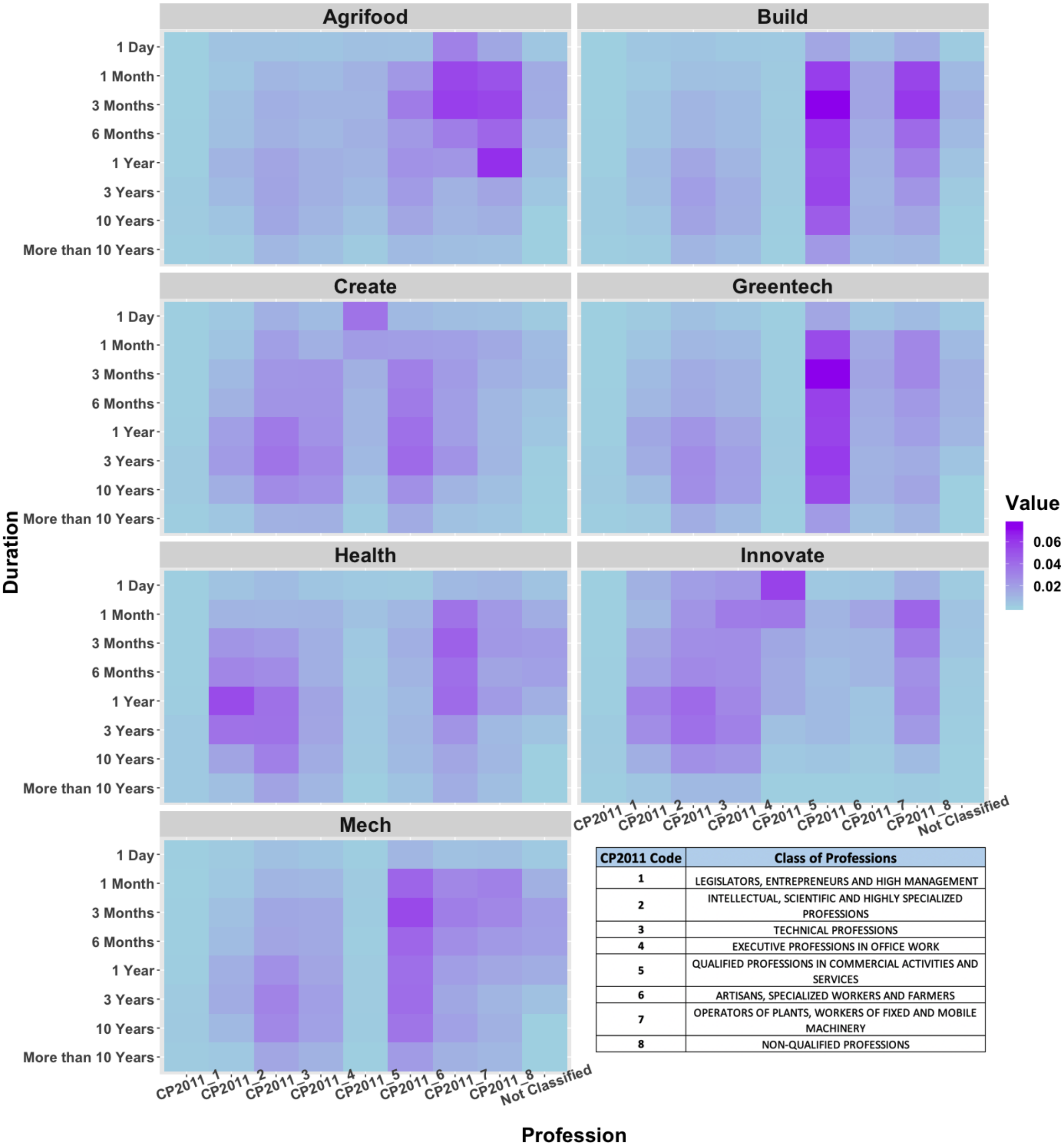


Figure 38: Correlation between ISTAT professional class and duration of contract by sector

To summarise, the decade under examination was a period of great structural and technological transformation. There is reason to believe that the process of digitisation, as in other highly developed areas of the country, has had no adverse effects on employment levels (Paba et al, 2020). Moreover, techhnological innovation should be seen as an opportunity, finalizing it to support workers in their tasks, instead of replacing semi-skilled or unskilled jobs (Rodrik, 2020). The latter is ever more relevant in a world affected by a dramatic pandemic phenomenon, where technology has become ever more necessary (Acemoglu, 2020).
But the period analysed also came between the two worst crises that the productive system faced in the post-war period. It was also a decade in which competition from emerging and newly industrialised countries became extraordinarily intense. For both reasons, unsurprisingly, it became even more evident that firms are geared towards limiting employment and the extensive use of short-term employment contracts. Both instruments aim to streamline processes and, above all, to reduce production costs.
Particularly penalised, as one would have expected, were jobs in which the qualifications and levels of education were lower. In other words, product competition in the market is undergoing change, and this has inevitable repercussions – not all positive – on the labour markets. The flip side is that the increase in university degree holders among new hires is still a relatively limited process.

In the context that has rapidly taken shape, it is particularly important to understand how this set of factors has acted on the demand for new and old professions and on the emerging demand for trades and skills. The following two paragraphs are dedicated to this objective.

4.5 Winners and losers: professional profiles and skills

Using the methodology set out in the previous pages (paragraph 4.3.2), the trends in the demand for professional figures and skills during the decade under consideration are presented below. Some of the main results are detailed in the appendix. The trend for the balances is represented through the use of graphs which indicate, for each sector, the professional profiles that have increased and the professional profiles for which demand has fallen to a greater extent. The same analysis and representation have been used for skills.

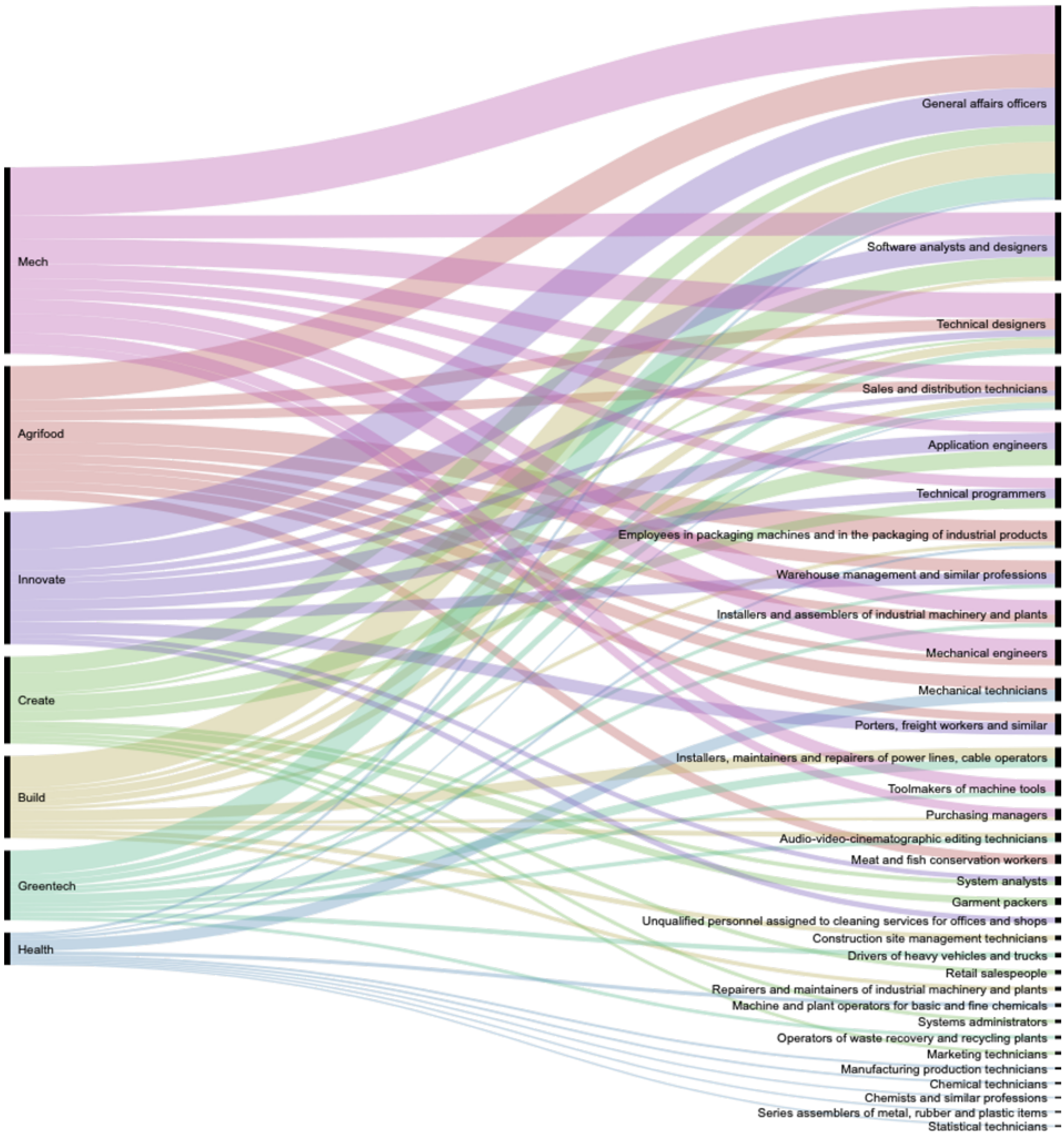


Figure 39: Professions with higher positive employment balance in the period 2008-2017

Some indications for reading the graphs follow. The occupational profiles and skills are ordered with respect to the cumulative values of the employment balance in the various sectors; the membership of one (or more) sectors can be obtained from each of the branches shown in the graphs. The thickness of the branch represents the size of the balance. The name given is that derived from the international classifications.

Figure 39 shows the only professions with highly positive employment balance for the period 2008-2017. In particular, the profiles with the highest employment balance were mainly associated with the Mech, Agrifood and Innovate sectors: the two sectors showing growth and the large mechanics and mechatronics sector which is widespread in the region. As is clear, demand is very diversified. Very briefly, (the reader should refer to Appendix 1 for more detailed data), we note that there are at least four types of professional profiles for which demand has clearly increased in recent years: 1. Technical profiles with medium- or high-level qualifications (software analysts and designers, technical designers, application engineers, technical programmers, installers and assemblers of industrial machinery and plants, chemists and similar professions etc.) 2. Management administrative profiles with a broad spectrum of skills (general affairs officers, system administrators, warehouse management and similar professions) 3. Sales and marketing staff (sales distribution technicians, marketing technicians). In addition to these: 4. A rather varied and significant component of workers with lower levels of qualifications (machine and plant operators, series assemblers, employees in packaging machines, drivers, garment packers, freight workers and similar, retail sales people, cleaning operators, porters etc.).

Overall, the picture is that of a productive system in which manufacturing is still of great importance and there is a strong differentiation between different manufacturing activities, and new professions are mixed with professions and trades which have been codified for decades, also in relation to relatively newer sectors.

New elements (and for which there is less evidence) emerge from the analysis of the demand for skills.

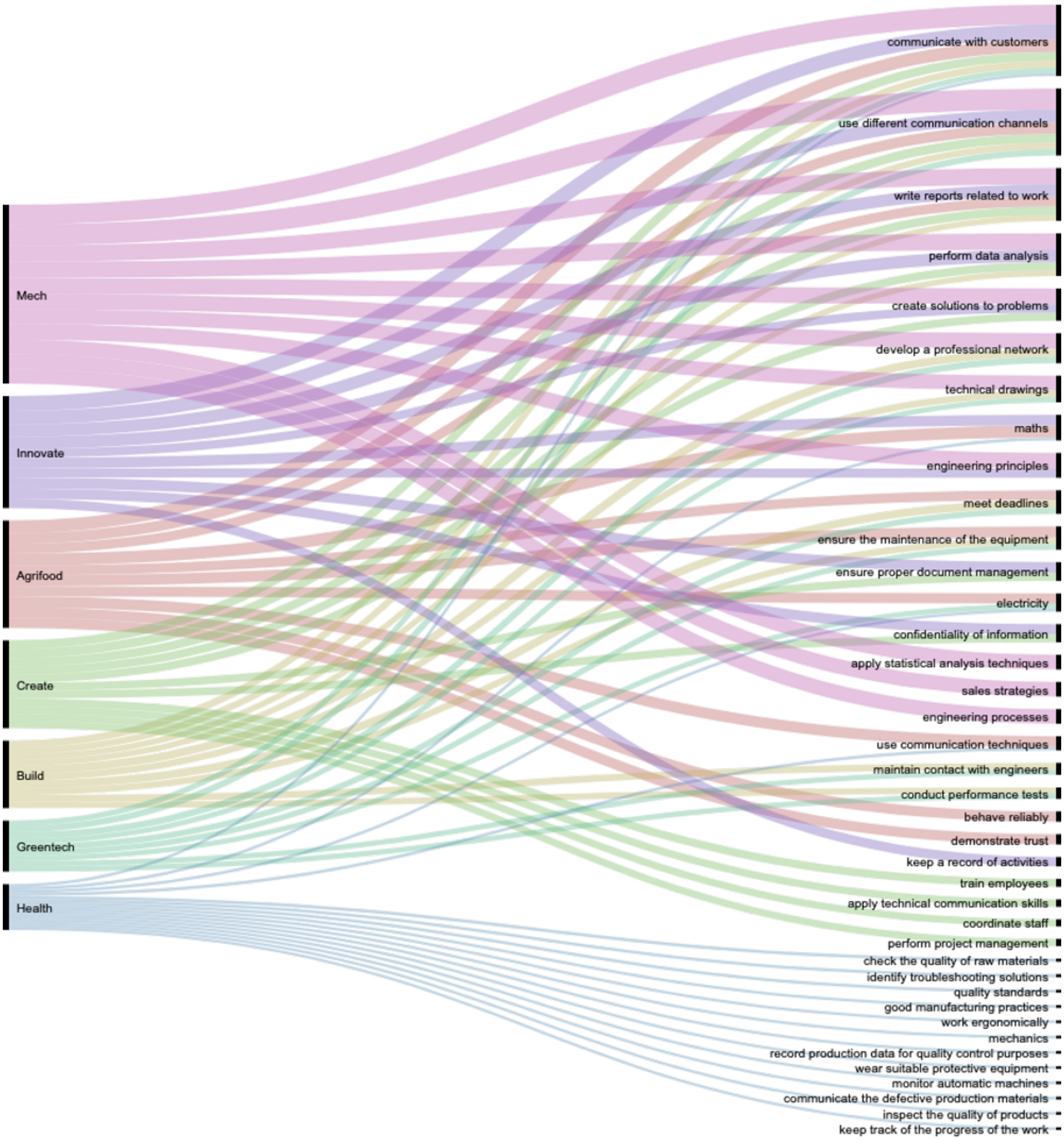


Figure 40: Extract of skills associated with profiles with higher positive employment balance in the period 2008-2017

Figure 40 shows the ranking of the 30 most requested skills and abilities (summary) over the decade by sector.

In line with our conclusions in the analysis of professional profiles, technical skills remain of considerable importance: from the ability to use CNC machines to the demand for data analysis skills, the latter evidently linked to the growing number of applications for computer science and digitisation of processes. However, the clearest data point to emerge is that the top of the ranking features skills ascribable to the notions of soft skills and cross-cutting: communication, problem-solving, co-ordination of working groups and staff, project management and timing, and so on, are thereby frequent. In this context, customer management skills (from communication to identifying their wishes etc.) are in particular demand at the company level. Although the Mech, Innovate and Agrifood sectors, given their relevance, continue to have a significant impact on the relevant profiles, the demand for these types of skills is increasing across all sectors.

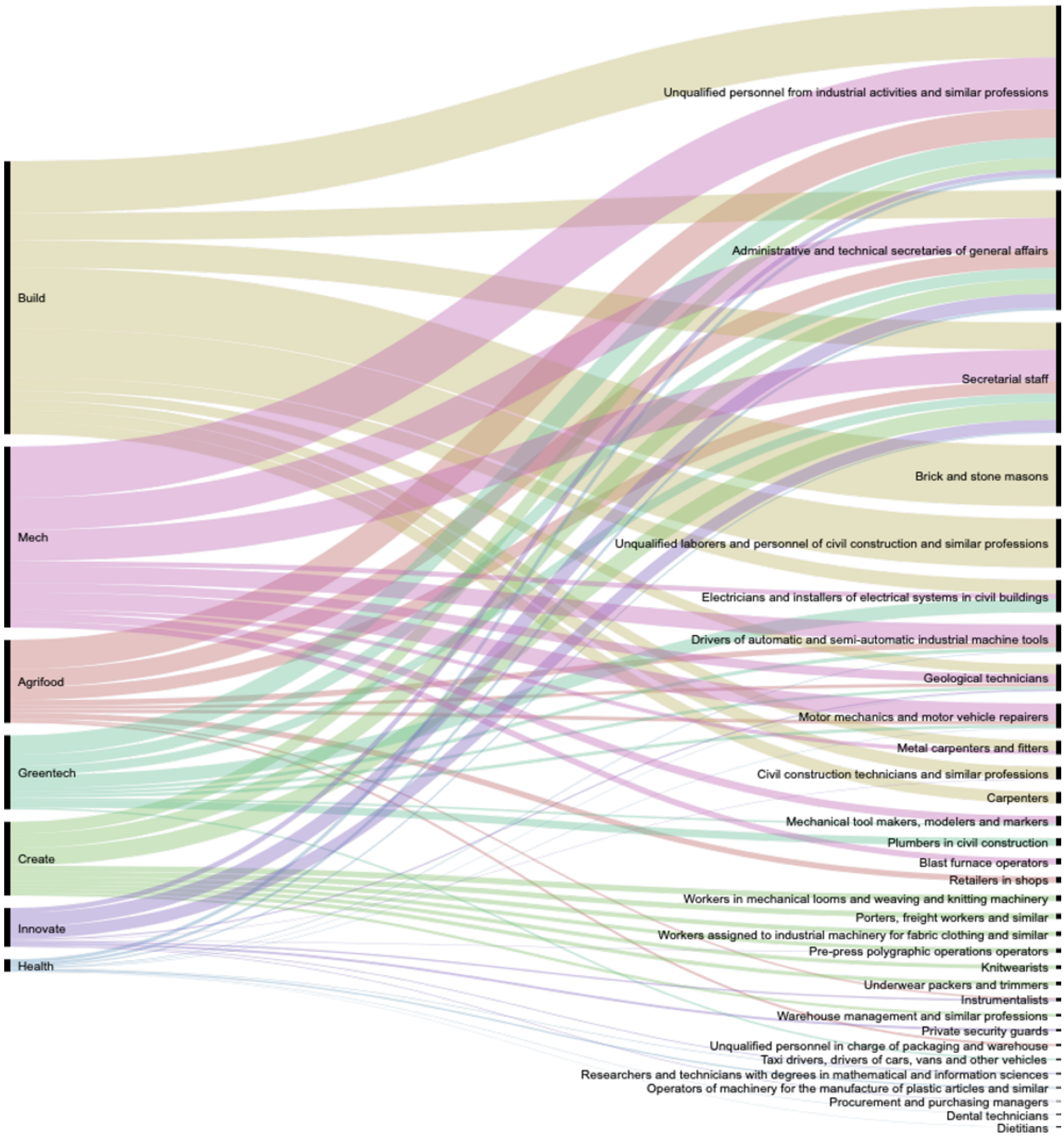


Figure 41: Professions with higher negative employment balance in the period 2008-2017

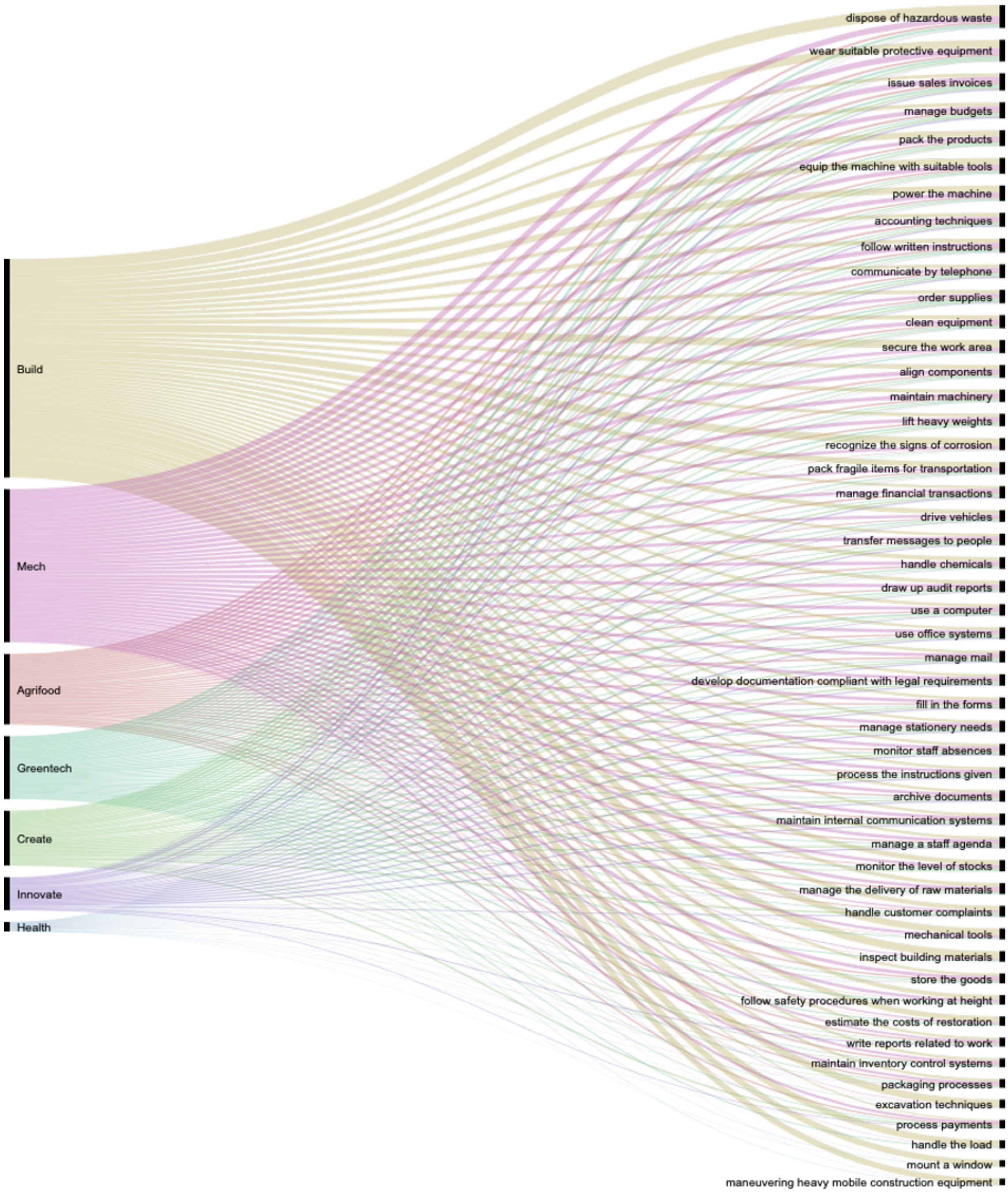


Figure 42: Extract of skills associated with profiles with higher negative employment balance in the period 2008-2017

The results for professional figures and skills with a larger negative employment balance are shown in Figure 41 and Figure 42. More detailed information is also

available for declining professional profiles in the Appendix (Appendix 2).
As with the professional profiles undergoing growth, the groups with declining professional profiles are also easily identified. Over the course of the decade, the contraction in demand affected essentially two groups: 1. Traditional trades (carpenters, masons etc.) 2. Some highly repetitive roles in production and offices. Both were concentrated in the Build and Mech sectors.
The traditional trades therefore saw an acceleration of a process that has been going on for decades during this period. For the latter, there is clearly a major process underway of changing and reorganising production processes and office work. Secretaries and secretarial workers, common to all sectors, are the roles that have suffered the largest cumulative negative impact.
The skills analysis confirms the observations for professional profiles. Being replaced are office tasks often related to information management, duties related to operational professional profiles (maintenance and cleaning of machinery, use of machine tools etc.) or associated with traditional trades. However, it should be noted that the size of the individual flows is very small in terms of absolute values: a sign of generalised adjustments, but which are in some sense marginal: whole blocks of skills are not disappearing, but there is, rather, a progressive process of replacing old skills with new ones. In this too, as in other times of great transformation, change is the result of a slow and continuous process, modulated by the system of industrial relations and by a variety of institutional factors. The interpretation of this result will be covered in the conclusions. Here, we will simply note that, over the course of the decade, it would seem to have affected the professions that are based on routine tasks and therefore subject to the highest risk of automation (Autor, 2015; Frey et al., 2017), and profiles with a high intensity of manual work also variously replaced by technology, organisational change and modular production. This brings us to the matter of the specific impact of the new digital paradigms associated with Industry 4.0.

4.6 The demand for digital skills

For the regional production system, as for other advanced economies, the Fourth Industrial Revolution, i.e. the automation and interconnection of production processes and the management of information flows, is perhaps the biggest challenge. The new flow involves production systems and processes, organisational dimensions, professional systems and, more generally, approaches to work. The new "intelligent factory" will have to control and manage production processes through the use of new digital and automated tools. The key technologies on which the technological revolution will be based concern areas such as cyber security, big data, cloud computing, augmented reality, robotics, rapid prototyping, radio frequency identification and tracking, super-connection of plants and 3D printing, but also new

approaches to work, process management and human resources management.
In the face of changes of this magnitude, trades, professions and, with them, the knowledge, skills and tasks required of the worker will change, triggering both adaptation and – even substantial – modification of individual jobs and work positions. The main players in these processes are, on the one hand, companies, who not only acquire the new profiles but also contribute to training them, and the whole training system: schools of all levels and universities. Introducing new machinery is not a sufficient condition for adapting organisations and businesses to new technologies and to changing product needs and demand. Properly carrying out an activity requires not only the idiosyncratic knowledge and skills typical of new technologies, but also the ability to implement them. In the face of the profound changes induced by digitisation, the efficiency of the enterprise and its adaptability, as on other occasions of great change, requires adequate technology, organisation and skills.
From this point of view, both for the scholar and for the public operator, one of the steps necessary for greater understanding of the current transformation of production systems is being able to outline the scope of industry 4.0 skills and define shared semantics. The tool used is the dictionary enhanced with technologies 4.0 from Chiarello et al. (2018) referred to in the preceding paragraphs. The enriched dictionary is the result of data mining work which started with the definition of a "seed list" of technologies 4.0 collected from scientific publications and papers. Once the manual review of the list was performed, the latter was enlarged using Wikipedia, collecting all the hyperlinks related to the seed elements. The output of the previous workflow was an enriched dictionary of 1211 technologies and more than 30000 relationships among them. Through the connections deduced from the links between the Wikipedia pages, the technologies were represented in a graph and automatically clustered by an algorithm in homogeneous groups. More specifically, the dictionary contains lists of regular expressions[34] which is a sequence of characters that define a search pattern. The patterns are used by string searching algorithms for "finding" or "finding and replacing" operations on strings, or for input validation, and it allowed the tool to capture the different orthographic declinations in which a technology could be written. In the present work, the analyses were performed with Software R. The dictionary was used to identify the ESCO skills that contain at least on technology and, for this reason, the ones that could be categorized as 4.0 (even if it may represent a light stretch). The analysed skills were the only ones related to the professional profiles most requested by companies, and it allows the creation of a ranking in terms of their relevance in the regional production chains. The authors deliberately filtered the results that seem less valuable, just like the term "computer" and the ones strictly related to its components ("software", "hardware", and so on), that are ever more considered mandatory and

[34] A regular expression, regex or regexp (sometimes called a rational expression)

findable everywhere (Fareri et al., 2020).

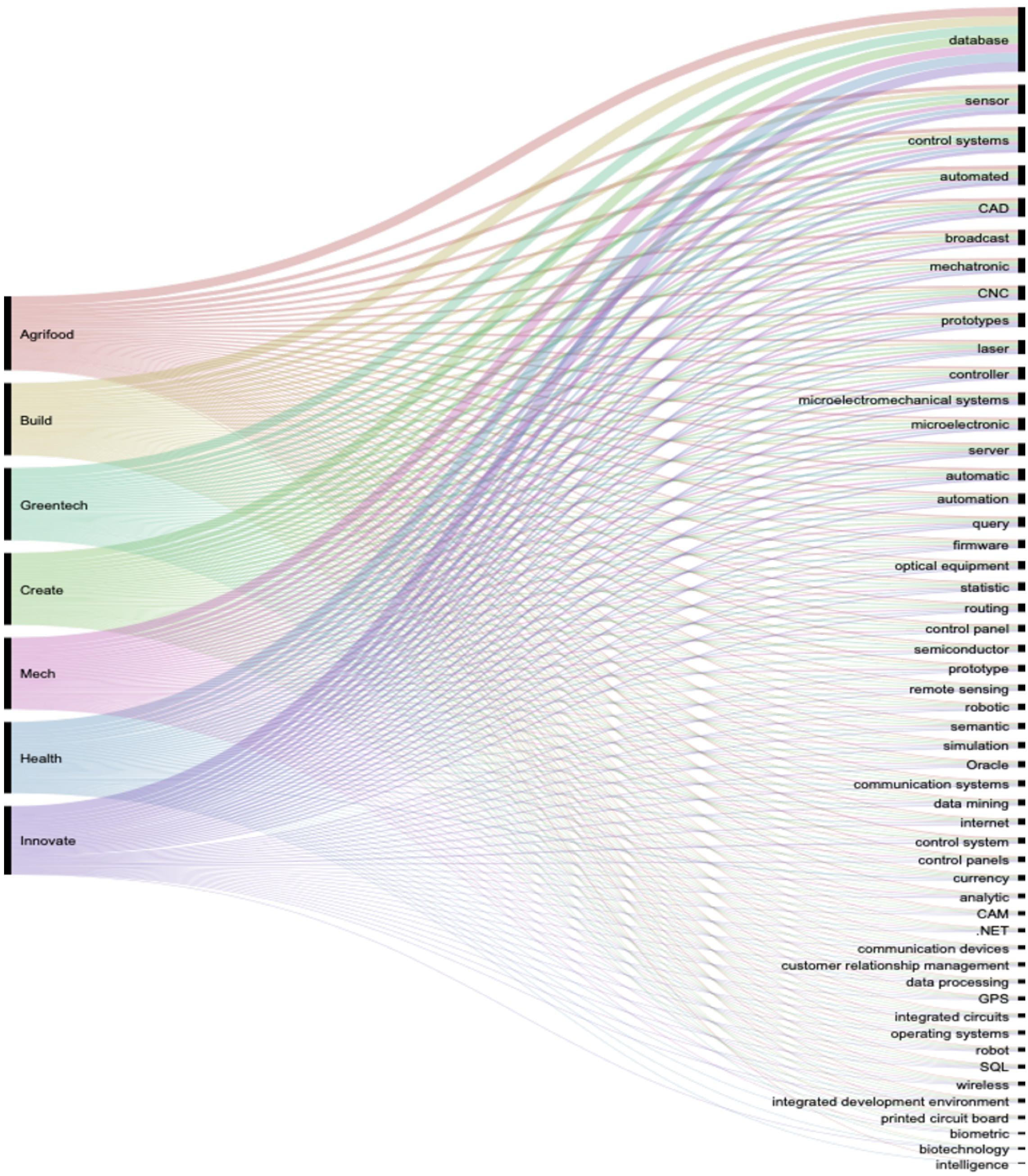


Figure 43: Technologies associated with professional profiles with positive employment balance

Figure 43 shows the results of the projection of the 4.0 dictionary on only oc-

cupational profiles with a positive employment balance. As the figure shows, the association of 4.0 technologies and skills in professional profiles undergoing growth reveals a very complex set of skills and abilities that concerns a wide range of business functions and production processes.
The relevance ranking, derived from the position in the right-hand column, is calculated with reference to the number of times that a given 4.0 technology appears in a descriptive profile declaration or in a skill associated with the professional profiles[35].
In relative terms, skills related to the creation and management of databases (and therefore to the collection and structuring of "data" in a broad sense) are particularly important for all the sectors. This is followed by sensor technology and signal transmission, crucial from an intelligent factory and Internet Of Things (IOT) point of view; process automation; CAD and simulation technologies.

One of the basic data is the cross-cutting nature of these technologies/skills between the different sectors.
This result is widely expected but nevertheless important. One of the distinctive aspects of the Fourth Industrial Revolution, compared to the previous ones (in addition to the automation of complex functions) is the transition from "sectoral" artificial intelligence to a general one which is widespread in the various economic activities (Bianchi, 2017).

A second exercise to assess the specific impact of individual technologies consisted of combining them into similar groups (clusters), then calculating the concentration of the new technologies present in the individual sectors with respect to the total. Fourteen clusters were thus obtained, entirely covering the spectrum of enabling technologies outlined by the Boston Consulting Group (2015): The result is shown in Figure 44. In this family of technologies there is a predominance of those belonging to the cloud computing and remote data management cluster; also significant is the presence of simulation technologies, and experimental analysis tools aimed at evaluating and predicting the dynamic development of a series of events under specified constraints, typically construction of scenarios, expected demand estimates in particular markets etc. (Fantoni et al., 2017).

[35] The dictionary also captures skills that straddle the digital and non-digital. In order to avoid assigning importance to non-digital skills, a threshold of ten recurrences has been imposed as a convention. Below this threshold, indeed, there are skills such as computer use which are now present in any job description.

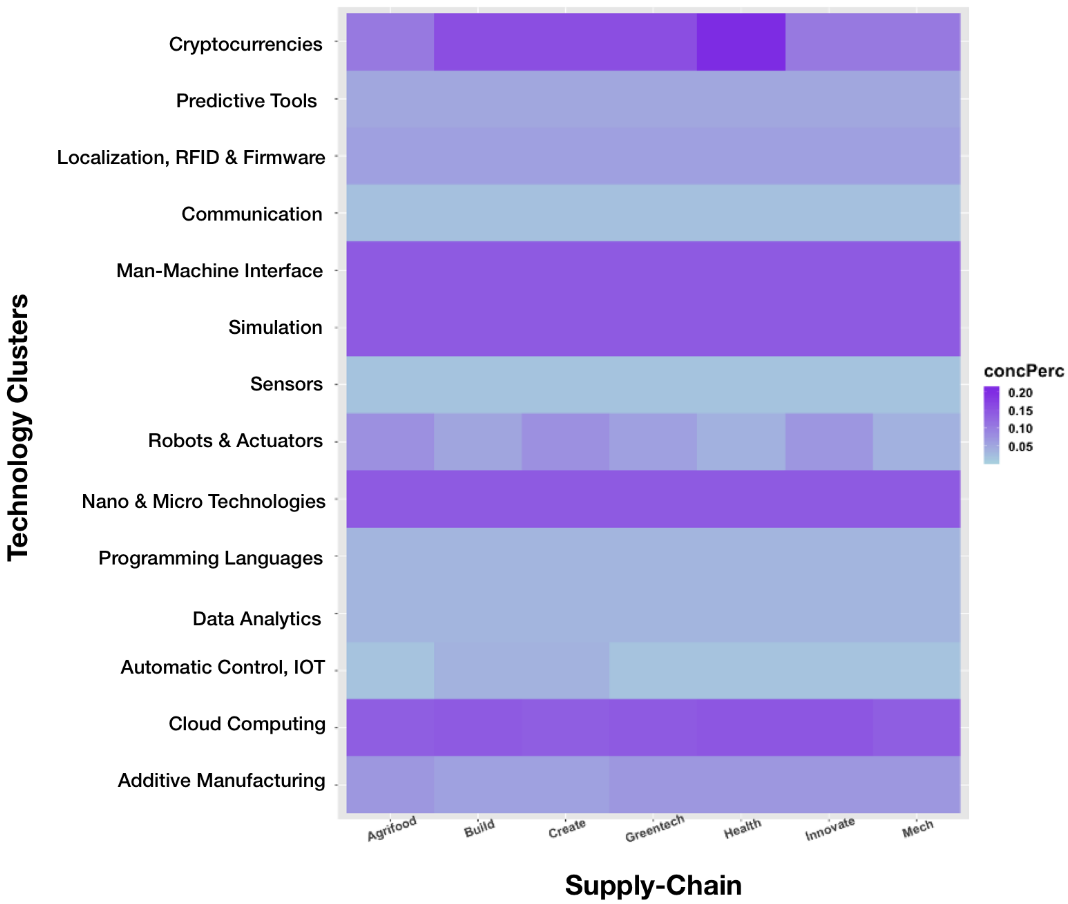


Figure 44: Impact of technology clusters on different sectors – Percentage values

With regard to blockchain technology and crypto-currencies (meaning a representation of the value based on cryptography), these have an incidence in all sectors, particularly Health. There are two possible reasons for this: centralised management of medical expenses and the management of confidential data that require transferability. In general, healthcare blockchains ensure that different actors share access to their networks without compromising the security and integrity of business data. Data analytics, programming languages (preparatory) and predictive tools follow at more modest concentrations. The results obtained are in line with other surveys on the adequacy and digital maturity of Emilian companies (Fareri et al., 2019).

	Agrifood	Build	Mech	Health	Create	Innovate	Greentech	Total
Total Hired	648.437	783.994	698.667	60.188	280.185	340.077	256.124	3.067.672
Not Digitized Hired	295.527	252.171	134.715	10.167	47.682	123.495	58.207	921.964
Digitized Hired	352.910	531.823	563.952	50.021	232.503	216.582	197.917	2.145.708
% Digitized Hired	54	68	81	83	83	64	77	70
% Not Digitized Hired	46	32	19	17	17	36	23	30

Table 25: Hires with digital skills out of the total number of hires for the period 2008-2017 – Absolute values and percentage values

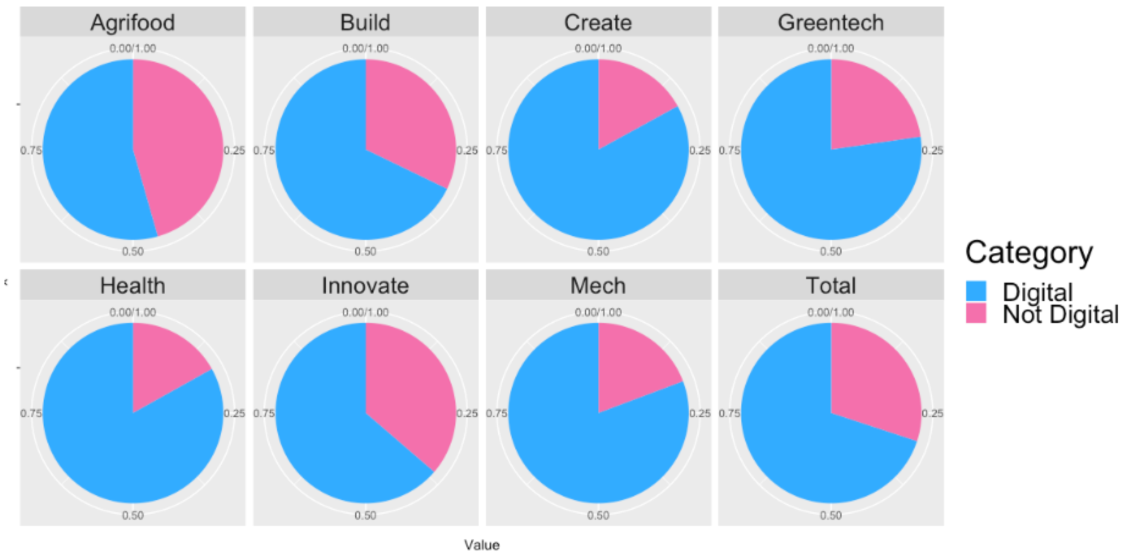


Figure 45: Hires with digital skills out of the total number of hires for the period 2008-2017 – Percentages

The last datum on which we would like to focus is the distribution of digital skills in Emilia-Romagna. Using the analysis tool previously presented, a simple measure of the distribution of digital skills was constructed using the ratio of hires with digital skills to the total number of employees hired during the period under consideration. The results are given in Table 25 and Figure 45. They show that, despite some variability, the vast majority of working relationships created over the course of the decade in all sectors required digital skills. This is as if to say that the Fourth Industrial Revolution is on the move leaving very clear signs in most professional profiles.

4.7 Summary and conclusions

The previous pages detailed a survey of the labour markets in Emilia-Romagna based on data deriving from mandatory notifications provided by companies for the decade 2008-2017, analysed starting out from the production chains, units of analysis and implementation of policies adopted by the Smart Specialisation Strategy.

This paper is divided into two parts.
The first involves an analysis of the size of the different sectors in the regional economy and an analysis of the incoming and outgoing movements generated by the companies in each sector during the course of the decade. The results provide insights into the composition and development of the different sub-markets. Beyond a clear identification of the sectors which have contracted and those which have developed– something which can be deduced from many other databases– flow analysis allows us to highlight two main elements. The first, which is widely known, consists of the spread, also in Emilia-Romagna, of work relationships of a short or very short duration, across all the sectors. Many of the Emilian labour markets show a typically dual structure with increasing proportions of new hires with markedly lesser safeguards and protections than older workers. The second fact is that all evidence shows, even in years when we observe a sea change in the production systems for which many of the constituent coordinates are being redesigned, the level of tertiary education of new hires continues to be low in relation to the trends typical of other advanced economies. This does not mean that sources capable of producing advanced knowledge are drying up in the industrial fabric. It certainly has repercussions on the capacity to differentiate and reallocate resources to emerging sectors and sectors of higher added value, however. Both raise issues of great importance for policy makers.
In relation to the first theme, it is sufficient to observe that short durations, starting from the seminal works of Jacob Mincer and Gary Becker, are the greatest obstacle to the development of human capital. In particular, considering job training, short durations determine an inefficient result: the company does not invest in the worker and the worker has no incentives to invest in the company. In the Italian case, a vast literature has highlighted that the fragmentation of careers for new entrants is one of the factors that determine unemployment, exit from the labour force and, in the long term, social inequality (Contini and Leombruni, 2006; Addabbo and Solinas, 2012; Contini, 2019; Mingione, 2020).
Equally important consequences derive from the results on the educational levels of the employees and new entrants. Here we do not want to reiterate the well-known and widely discussed data relating to the insufficient training of new hires, and in particular to the low number of graduates over total employment, but a somewhat complementary feature. Underlining, as has been done previously, that the one

studied is still an economy of high school graduates/industrial graduates and not of engineers, we want to remind you that in productive systems such as those examined in these pages, even in periods of radical transformation, regional governments must continue to provide great attention and great care, as happened in the past, to secondary education and, in particular, to technical and professional schools (Brusco, 1982; Rinaldi, 2005).

A new element emerges in the second part of the paper: the characteristic aspects of the analysis of flows are indeed reconnected with the emerging demand for professional profiles and skills. With regard to professional profiles and skills, a methodology is proposed which is capable of providing a measure, built on employment balances, between incoming flows and outgoing flows. This measure makes it possible to identify the winners and losers in a small economy open and exposed to all the changes taking place in the world economy such as the Emilian one. From this point of view, the focus is shifting towards the processes of replacing employment profiles and skills induced in particular by the spread of new technologies. Without revisiting the points made previously, we note that the results obtained confirm an important part of what the literature about the effects of new technologies on the labour markets tells us. Along with a downsizing of the old trades, in which the characteristic element is physical effort and manual skill – a process that has been ongoing for some decades – the jobs that appear most at risk of replacement are routine factory line and office jobs. This is in keeping with the findings of seminal studies on the effects of digital technologies, whether they are based on professions (Frey et al., 2017) or on the tasks and skills associated with professional profiles (Autor. 2015).
From this point of view, which is not refuted by the empirical analysis developed in the essay, we are not arguing that "dropping occupations" are totally automatable, and we don't believe they are expected to disappear. As also highlighted in the McKinsey Report (2017), full automation is a process whose methods and implementation times are subject to the costs of implementing new technologies, their degree of social acceptance, labour market regulation systems, the effectiveness of training systems in retraining, and reconverting workers to the new conditions (Nedelkoska et al., 2018).
Moreover, the top positions of the "winning" skills ranking are held by soft skills: communication, problem-solving, coordination of working groups and staff, project management and timing (and so on), are the most frequent. The result is very interesting and in accordance with literature; in fact, many studies stressed the advantage of acquiring soft skills in order to face the digital wave (Chryssolouris et al., 2013; Gorecky et al., 2014; Weber, 2016), outlining that a skill gap in this context will have a negative impact on the workforce (Frank et al., 2019; Bauer et al., 2011; Bridgstock, 2011; Dobrunz et al., 2006; Haukka, 2011; Cooper and Tang, 2010). It

is also important to outline that the belief that they are innate characteristics of individuals is disappearing and many scholars have worked in the last years searching for innovative ways to facilitate their development (Sanz et al., 2019; Tseng et al., 2019; Duran-Novoa et al., 2011), their assessment (Bohlouli, et al., 2017), their comprehension (Chechurin et al., 2016) or more specifically identifying how they impact digital jobs (Hendon et al., 2017). For what concerns technical skills, and more specifically digital skills, there is as a clear request of those related to data management which is considered a priority for actually beginning the change (Roy, 2016; Colegrove, 2017). Furthermore, there is a predominance of cloud computing and experimental analysis tools aimed at evaluating and predicting the dynamic development of events under specified constraints, also known as Simulation technologies, that are ever more requested by firms in Emilia-Romagna (Solinas et al., 2019) and deeply described by (Fantoni et al., 2017).

The final part of the essay outlines a method for estimating how much labour demand is specifically associated with 4.0 technologies and which of them are now most in demand by companies. It shows that digital skills and knowledge are required of three out of four new hires. Against this background, the fact which stands out is that the management of information flows is by far the most characteristic element of the ongoing transformation processes. This element emerges strongly in all the production chains, both the new chains and those more typical of Emilian manufacturing tradition. The management of information flows is the heart of the "intelligent factory", the glue holding together new and old industrial knowledge, new and old crafts, and the ability to learn and the ability to do.

The research presented in this paper is to some extent exploratory, and the proposed methodology must certainly be refined. We believe, however, that it represents a useful tool for analysing labour markets, changes in the industrial structure and for the construction of informed and aware economic policies itself.

4.8 Appendix 1 - Professions with the most Positive Employment Balance [2008-2017]

Sector	Employment Balance (+)	Job Profile CP2011
Mech	4578	General affairs officers
Innovate	3551	General affairs officers
Agrifood	3214	General affairs officers
Build	3001	General affairs officers
Mech	2389	Technical designers
Greentech	2279	General affairs officers
Mech	2172	Software analysts and designers
Innovate	2035	Software analysts and designers
Agrifood	2005	Employees in packaging machines and in the packaging of industrial products
Create	1880	Software analysts and designers
Mech	1820	Mechanical engineers
Innovate	1587	Application engineers
Create	1571	General affairs officers
Create	1509	Application engineers
Mech	1383	Sales and distribution technicians
Mech	1361	Installers and assemblers of industrial machinery and plants
Agrifood	1201	Warehouse management and similar professions
Innovate	1181	Porters, freight workers and similar
Mech	1153	Toolmakers of machine tools
Agrifood	1123	Mechanical technicians
Health	1095	Mechanical technicians
Agrifood	1079	Technical designers
Innovate	1049	Warehouse management and similar professions
Build	1039	Installers, maintainers and repairers of power lines, cable operators
Mech	1001	Application engineers
Mech	990	Technical programmers
Innovate	974	Technical programmers
Create	906	Technical programmers
Agrifood	888	Sales and distribution technicians
Agrifood	882	Meat and fish conservation workers
Greentech	865	Installers, maintainers and repairers of power lines, cable operators
Agrifood	827	Porters, freight workers and similar
Mech	820	Purchasing managers
Build	814	Technical designers
Agrifood	801	Installers and assemblers of industrial machinery and plants
Create	680	Garment packers
Innovate	643	Technical designers

Sector	Employment Balance (+)	Job Profile CP2011
Agrifood	637	Mechanical engineers
Build	601	Sales and distribution technicians
Greentech	586	Technical designers
Greentech	572	Sales and distribution technicians
Innovate	537	Sales and distribution technicians
Innovate	530	Unqualified personnel assigned to cleaning services for offices and shops
Build	478	Audio-video-cinematographic editing technicians
Build	464	Construction site management technicians
Innovate	457	System analysts
Greentech	455	Drivers of heavy vehicles and trucks
Greentech	409	Audio-video-cinematographic editing technicians
Create	402	Retail salespeople
Create	398	System analysts
Greentech	394	Toolmakers of machine tools
Greentech	385	Installers and assemblers of industrial machinery and plants
Build	384	Software analysts and designers
Build	381	Repairers and maintainers of industrial machinery and plants
Health	363	Machine and plant operators for basic and fine chemicals
Build	345	Employees in packaging machines and in the packaging of industrial products
Create	342	Systems administrators
Greentech	333	Warehouse management and similar professions
Greentech	329	Operators of waste recovery and recycling plants
Create	303	Marketing technicians
Build	283	Purchasing managers
Health	259	General affairs officers
Create	259	Technical designers
Health	256	Employees in packaging machines and in the packaging of industrial products
Health	225	Manufacturing production technicians
Health	209	Chemical technicians
Health	172	Chemists and similar professions
Health	170	Series assemblers of metal, rubber and plastic items
Health	164	Statistical technicians
Health	145	Sales and distribution technicians

4.9 Appendix 2 - Professions with the most Negative Employment Balance [2008-2017]

Sector	Employment Balance (-)	Job Profile CP2011
Build	9277	Brick and stone masons
Build	7896	Unqualified personnel from industrial activities and similar professions
Mech	7800	Unqualified personnel from industrial activities and similar professions
Build	7431	Unqualified laborers and personnel of civil construction and similar professions
Mech	4902	Administrative and technical secretaries of general affairs
Mech	4719	Secretarial staff
Agrifood	4474	Unqualified personnel from industrial activities and similar professions
Build	4146	Administrative and technical secretaries of general affairs
Build	4132	Secretarial staff
Greentech	2966	Unqualified personnel from industrial activities and similar professions
Agrifood	2723	Administrative and technical secretaries of general affairs
Create	2624	Secretarial staff
Mech	2619	Motor mechanics and motor vehicle repairers
Mech	2601	Drivers of automatic and semi-automatic industrial machine tools
Create	2218	Administrative and technical secretaries of general affairs
Innovate	2101	Administrative and technical secretaries of general affairs
Build	2082	Electricians and installers of electrical systems in civil buildings
Greentech	2030	Electricians and installers of electrical systems in civil buildings
Agrifood	1945	Secretarial staff
Build	1842	Civil construction technicians and similar professions
Create	1814	Unqualified personnel from industrial activities and similar professions
Innovate	1813	Secretarial staff
Build	1770	Carpenters
Greentech	1723	Administrative and technical secretaries of general affairs
Build	1479	Metal carpenters and fitters
Build	1466	Geological technicians
Mech	1400	Geological technicians
Greentech	1354	Secretarial staff
Mech	1196	Mechanical tool makers, modelers and markers
Greentech	1176	Plumbers in civil construction
Mech	923	Blast furnace operators
Agrifood	903	Retailers in shops
Create	888	Workers in mechanical looms and weaving and knitting machinery
Agrifood	882	Drivers of automatic and semi-automatic industrial machine tools

Sector	Employment Balance (-)	Job Profile CP2011
Create	817	Porters, freight workers and similar
Mech	766	Electricians and installers of electrical systems in civil buildings
Innovate	754	Unqualified personnel from industrial activities and similar professions
Create	700	Workers assigned to industrial machinery for fabric clothing and similar
Mech	605	Metal carpenters and fitters
Create	588	Knitwearists
Create	582	Underwear packers and trimmers
Create	576	Pre-press polygraphic operations operators
Agrifood	556	Motor mechanics and motor vehicle repairers
Health	544	Unqualified personnel from industrial activities and similar professions
Greentech	538	Drivers of automatic and semi-automatic industrial machine tools
Agrifood	505	Geological technicians
Greentech	472	Motor mechanics and motor vehicle repairers
Greentech	444	Geological technicians
Create	413	Warehouse management and similar professions
Health	400	Administrative and technical secretaries of general affairs
Innovate	361	Private security guards
Agrifood	353	Unqualified personnel in charge of packaging and warehouse
Greentech	285	Mechanical tool makers, modelers and markers
Agrifood	269	Instrumentalists
Greentech	266	Taxi drivers, drivers of cars, vans and other vehicles
Innovate	255	Instrumentalists
Health	227	Operators of machinery for the manufacture of plastic articles and similar
Innovate	207	Geological technicians
Health	202	Secretarial staff
Innovate	147	Researchers and technicians with degrees in mathematical and information sciences
Innovate	125	Procurement and purchasing managers
Health	122	Drivers of automatic and semi-automatic industrial machine tools
Health	104	Researchers and technicians with degrees in mathematical and information sciences
Health	81	Motor mechanics and motor vehicle repairers
Innovate	78	Pre-press polygraphic operations operators
Health	75	Dental technicians
Innovate	74	Civil construction technicians and similar professions
Health	68	Geological technicians
Health	51	Dietitians

4.10 References

Abbott, A., (1993). *The sociology of work and occupations.* Annu. Rev. Sociol. 19 (August), 187–209.

Acemoglu, D., Autor, D., (2010). *Skills, tasks and technologies: implications for employment and earnings.* doi:10.3386/w16082

Acemoglu, D., Restrepo, P. (2020). *The wrong kind of AI? artificial intelligence and the future of labour demand.* Cambridge Journal of Regions, Economy and Society, 13 (1), pp. 25-35.

Acemoglu A., (2020). *Remaking the Post-COVID World.* Sixth Richard Goode Lecture, IMF

Addabbo T. Solinas G., eds., (2012). *Non-standard employment and quality of work. The case of Italy,.* AIEL series, vol. III, Heidelberg, Physica-Verlag.

Ajello, A.M., (2002). *La competenza*, Bologna, Il Mulino.

Alabdulkareem, A., Frank, M. R., Sun, L., Alshebli, B., Hidalgo, C., Rahwan, I., (2018). *Unpacking the polarization of workplace skills.* Science Advances,4(7). doi:10.1126/sciadv.aao6030

Alfonso-Hermelo, D., Langlais, P., Bourg, L., (2019). *Automatically learning a human-hesource hntology from professional social-network data.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11489 LNAI, pp. 132-145. DOI: 10.1007/978-3-030-18305-9_11

Antonelli, G., Nosvelli, M., (2008). *Demand for skilled labour services, job design and the 'revealed' learning function.* 10.4324/9780203937396.ch4.

Arntz, M., Gregory, T., Zierahn, U., (2017). *Revisiting the risk of automation.* Economics Letters,159, 157-160. doi:10.1016/j.econlet.2017.07.001

Autiero, G., Bruno B., Parrella, M. (2020). *Contenuto di istruzione della domanda di lavoro in Italia.*

Autor, D., (2015). *Why are there still so many jobs? The history and future of workplace automation.* Journal of Economic Perspectives. 29. 3-30. 10.1257/jep.29.3.3.

Autor, D., Dorn, D., (2009). *The growth of low skill service jobs and the polarization of the U.S. Labor Market.* doi:10.3386/w15150

Barney, J. B., (1991). *Firm resources and sustained competitive advantage.* Journal of Management, 17: 99-120.

Batazzi, M., (2015). *Le imprese toscane, fabbisogni professionali e competenze per il 2015.* 10.13140/RG.2.1.2941.8643.

Bauer, C., Viola, K., Strauss, C., (2011). *Management skills for artists: 'learning by doing'?* Int. J. Cult. Policy 17 (5), 626–644.

Becker, B., Gerhart, B., (1996). *The impact of human resource management on organizational performance: Progress and prospects.* Academy of Management Journal, 39: 779-801.

Benadusi, L., Molina, S. (2018), *Le competenze, Una mappa per orientarsi.* Bologna: Il Mulino.

Bianchi, P., (2017). *4.0 La nuova rivoluzione Industriale.* Il Mulino, Ed. 267

Bohlouli, M., Mittas, N., Kakarontzas, G., Theodosiou, T., Angelis, L., Fathi, M., (2017). *Competence assessment as an expert system for human resource management: a mathematical approach.* Expert Systems with Applications, 70, pp. 83-102. DOI: 10.1016/j.eswa.2016.10.046

Boselli, R., Cesarini, M., Mercorio, F., Mezzanzanica, M., (2018). *Classifying online job advertisements through machine learning.* Future Generation Computer Systems,86, 319-328. doi:10.1016/j.future.2018.03.035

Boston Consulting Group, (2015). *Industry 4.0: the future of productivity and growth in manufacturing industries.*

Bridgstock, R., (2011). *Skills for creative industries graduate success.* Educt. Training 53 (1), 9–26.

Brusco, S. (1982). *The Emilian Model: productive decentralization and social integration.* Cambridge Journal of Economics, vol. VI, n. 2, pp. 167-184.

Caruso, L., (2017). *Digital innovation and the fourth industrial revolution:*

epochal social changes? Ai Society, 33(3), 379-392. doi:10.1007/s00146-017-0736-1

Chechurin, L., Borgianni, Y., (2016). *Understanding TRIZ through the review of top cited publications.* Computers in Industry, 82, pp. 119-134.

Chiarello, F., Trivelli, L., Bonaccorsi, A., Fantoni, G., (2017). *Extracting and mapping industry 4.0 technologies using wikipedia.* Computers in Industry, 100, 244-257.

Chomsky, N., (1965). *Aspects of Theory of syntax,.* Cambridge, MA, The Mit Press; trad. it. Saggi linguistici. La grammatica generativa trasformazionale, Torino, Boringhieri, 1970.

Chryssolouris, G., Mavrikios, D., Mourtzis, D., (2013). *Manufacturing systems: skills competencies for the future.* Procedia CIRP 7, 17–24.

Clifton, J., Glasmeier, A., Gray, M., (2020). *When machines think for us: The consequences for work and place.* Cambridge Journal of Regions, Economy and Society, 13 (1), pp. 3-23.

Colegrove, L., Peters, J., (2017). *Big data analytics skills - A revolution lacking in revolutionaries.* Sustainable Engineering Forum 2017 - Topical Conference at the 2017 AIChE Spring Meeting and 13th Global Congress on Process Safety, pp. 2-10.

Colombo, E., Mercorio, F., Mezzanzanica, M., (2019). *AI meets labor market: Exploring the link between automation and skills.* Information Economics and Policy, 47, pp. 27-37. DOI: 10.1016/j.infoecopol.2019.05.003

Contini B., ed., (2002). *Osservatorio sulla mobilità del lavoro in Italia.* Bologna, Il Mulino.

Contini B., (2019). *Dai lavori "usa e getta" al jobs act.* Roma, Aracne.

Contini B., (2005). *Mobilità del lavoro e dei posti di lavoro. Problemi di misura e questioni aperte.* in B. Contini e U. Trivellato (2005), pp. 115-142.

Contini B., Leombruni R. (2006), *From work to retirement: A tale of bumpy routes. Review of Political Economy,* vol. 18, n. 3, pp-359-378.

Contini B. Revelli R. (1997), *Gross flows vs. net flows in the labour market: what is there to be learned.* Labour economics, n.3.

Contini B. Trivellato U. (2005), *Eppur si muove: dinamiche e persistenze nel mercato del lavoro italiano,* Bologna, Il Mulino.

de Vries, G.J., Gentile, E., Miroudot, S., Wacker, K.M., (2020). *The rise of robots and the fall of routine jobs.* Labour Economics, 66, art. no. 101885.

Dobrunz, J., Schöppner, H., Wolfram, A., (2006). *Anforderungen an den Mediennachwuchs* — die media.net employability-studie. media.net berlinbrandenburg e.V, Berlin.

Dubar, C., (1991). *La socialisation. Construction des identités sociales et professionnelles,* Paris, Colin; trad. it. La socializzazione. Come si costruisce l'identità sociale, Bologna, Il Mulino, 2004.

Duran-Novoa, R., Leon-Rovira, N., Aguayo-Tellez, H., Said, D., (2011). *Inventive problem solving based on dialectical negation, using evolutionary algorithms and TRIZ heuristics.* Computers in Industry, 62 (4), pp. 437-445.

European Commission, (2020). https://ec.europa.eu/esco/portal/escopedia/Skill

Fantoni, G., Cervelli G., Pira S., Trivelli L., (2017). *Industria 4.0 senza slogan*

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.

Fernández-Sanz, L., Gómez-Pérez, J., Castillo-Martínez, A., (2017). *E-Skills Match: A framework for mapping and integrating the main skills, knowledge and competence standards and models for ICT occupations.* Computer Standards Interfaces,51, 30-42. doi:10.1016/j.csi.2016.11.004

Fonseca, T., Lima, F., Pereira, S.C., (2018). *Job polarization, technological change and routinization: Evidence for Portugal.* Labour Economics, Volume 51, Pages 317-339.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Rahwan, I. (2019). *Toward understanding the impact of artificial intelligence on labor.* Proceedings of the National Academy of Sciences, 116(14), 6531–6539. doi: 10.1073/pnas.1900949116

Freddi, D., (2017). *Digitalisation and employment in manufacturing.* Ai Society, 33(3), 393-403. doi:10.1007/s00146-017-0740-5

Frey, C.B., Osborne, M.A., (2017). *The future of employment: How susceptible are jobs to computerization?* Technological Forecasting and Social Change, Sept 1–72.

Fulmer, I. S., Ployhart, R. E. (2013). *Our most important asset. Journal of Management,* 40(1), 161–192. doi:10.1177/0149206313511271

Galati, F., Bigliardi, B., (2019). *Industry 4.0: Emerging themes and future research avenues using a text mining approach.* Computers in Industry, 109, pp. 100-113.

Gentili, A., Compagnucci, F., Gallegati, M., Valentini, E., (2020). *Are machines stealing our jobs?* Cambridge Journal of Regions, Economy and Society, 13 (1), pp. 153-173.

Gorecky, D., Schmitt, M., Loskyll, M., Zühlke, D., (2014). *Human-machine-interaction in the Industry 4.0 era.* In: In Proceedings of the 2014 12th IEEE International Conference on Industrial Informatics (INDIN), Porto Alegre, Brazil.

Grillitsch M., Asheim B.T and Trippl M., (2018). *Unrelated knowledge combinations: the unexplored potential for regional industrial path development.* Cambridge Journal of Regions, Economy and Society

Grugulis, I., Vincent, S., (2009). *Whose skill is it anyway? 'Soft' skills and polarization.* Work Employ. Soc. 23 (4), 597–615.

Haltiwanger J., Manser M. Topel R., eds., (1999) *Labor Statistics Measurement Issues.* Chicago: University of Chicago Press.

Haukka, S., (2011). *Education-to-work transitions of aspiring creatives.* Cultural Trends 20 (1), 41–64.

Hendon, M., Powell, L., Wimmer, H. (2017). *Emotional intelligence and communication levels in information technology professionals.* Computers in Human Behavior, 71, 165–171. doi: 10.1016/j.chb.2017.01.048

ISTAT, *La classificazione delle professioni* (2013).

Jonnaert, Ph., (2009). *Compétences et socioconstructivisme, .* Un cadre théorique, Bruxelles, De Boeck.

Karakatsanis, I., Alkhader, W., Maccrory, F., Alibasic, A., Omar, M. A., Aung, Z., Woon, W. L., (2017). *Data mining approach to monitoring the requirements of the job market: A case study.* Information Systems,65, 1-6. doi:10.1016/j.is.2016.10.009

Lado, A. A., Wilson, M. C., (1994). *Human resource systems and sustained competitive advantage: A competency- based perspective.* Academy of Management Review, 19: 699-727.

Lalé, E., (2020). *Loss of skill and labor market fluctuations.* Labour Economics, Volume 50, Pages 20-31.

Last, C., (2017). *Global commons in the global brain.* Technological Forecasting and Social Change, 114, 48-64.

Leigh, N.G., Kraft, B., Lee, H., (2020). *Robots, skill demand and manufacturing in US regional labour markets.* Cambridge Journal of Regions, Economy and Society, 13 (1), pp. 77-97

Levy, F., Murnane, R. (2004). *The new division of labor: how computers are creating the next job Market.* New York; Princeton; Oxford: Princeton University Press.

Lorentz, H., Töyli, J., Solakivi, T., Ojala, L., (2013). *Priorities and determinants for supply chain management skills development in manufacturing firms.* Supply Chain Manag. Int. J. 18 (4), 358–375

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.

McKinsey Global Institute, (2017). *A future that works: automation, employment, and productivity*, s.l.

Mingione E. (2020). *Lavoro: la grande trasformazione: l'impatto sociale del cambiamento del lavoro tra evoluzioni storiche e prospettive globali.* Milano, Feltrinelli.

Mirski, P., Bernsteiner, R., Radi, D., (2017). *Analytics in human resource management: the OpenSKIMR approach.* Procedia Computer Science,122, 727-734. doi:10.1016/j.procs.2017.11.430

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

Pontecorvo, C., Ajello, A.M. e Zucchermaglio, C., (1995). I *contesti sociali dell'apprendimento. Acquisire conoscenze a scuola, nel lavoro, nella vita quotidiana*, Milano, LED.

Pryima, S., Rogushina, J.V., Strokan, V., (2018). *Use of semantic technologies in the process of recognizing the outcomes of non-formal and informal learning*, CEUR Workshop Proceedings, 2139, pp. 226-235.

Rinaldi A. (2005), *The Emilian Model Revisited: Twenty years after*, Business History, vol. 47, n. 2, pp. 244-266.

Rodrik D., (2020). *Technology for All.* Project Syndacate

Roy, R., Stark, R., Tracht, K., Takata, S., Mori, M., (2016). *Continuous maintenance and the future – Foundations and technological challenges.* CIRP Annals - Manufacturing Technology, 65 (2), pp. 667-688. DOI: 10.1016/j.cirp.2016.06.006

Rosenberg M., (2009). *The surprising benefits of robots in the DC. Supply Demand Chain Executive*, 10 (2): 39-40

Rotman D., (2013). *How technology is destroying jobs.* MIT Technology Review, June 12. http://www.technologyreview.com/featuredstory/515926/how-technology-is-destroying-jobs

Rullani E., (2004). *Economia della Conoscenza.* Carocci editore

Sanz, I., Montero, J., Sevillano, X., Carrié, J.C., (2019). *Developing a videogame for learning signal processing and project management using project-oriented learning in ICT engineering degrees.* Computers in Human Behavior. 99.

Solinas G., (1990). *Alcune considerazioni sul turnover e le misure di durata dell'occupazione.* Politica Economica, anno VI, n. 3, pp. 449-80.

Solinas, G., Fareri S., Giordano V., (2019). *La maturità digitale delle imprese in Emilia-Romagna Primi risultati.* Working paper, Unioncamere e Regione Emilia-Romagna.

Tarry, A., (2019). *Coaching with careers and AI in mind: Grounding a hopeful and resourceful self fit for a digital world.* Abingdon, Oxon, England: Routledge.

Torelli N. Trivellato U. (1993), *Modelling inaccuracies in job-search duration data, Journal of Econometrics,* vol. 59, n.1-2, pp. 187-211.

Trivellato U. (2011). *Indagini dirette e fonti amministrative: dall'alternativa all'ancora compiuta integrazione,* Rivista di politica economica, n. 1, pp. 21-60.

Tseng, H., Yi, X., Yeh, H.-T. (2019). *Learning-related soft skills among online business students in higher education: grade level and managerial role differences in self-regulation, motivation, and social skill.* Computers in Human Behavior, 95, 179–186. doi: 10.1016/j.chb.2018.11.035

Van Laar, E., Deursen, A., Van Dijk, J., Haan, J. (2017). *The relation between 21st-century skills and digital skills: A systematic literature review.* Computers in Human Behavior. 72. 10.1016/j.chb.2017.03.010.

Weber, E., (2016). *Industry 4.0: job-producer or employment-destroyer?* Retrieved on March 2017, http://doku.iab.de/aktuell/2016/aktueller_bericht_1602.pdf.

Wilson R., (2013). *Skills anticipation - The future of work and education.*

World Economic Forum, (2016). *The Future of jobs.*

List of Figures

List of Tables