Università
Ca'Foscari
Venezia

Master's Degree

in Global Development and Entrepreneurship

Final Thesis

Modelling the innovativeness perception of businesses: an experimental analysis

Supervisor
Prof. Giacomo di Tollo

Graduand
Sara Ghilardi
Matriculation number 882943

Academic Year
2020 / 2021

Abstract

In this thesis we want to model businesses' perception of innovation on the basis of information contained in some components that are becoming a key-point in the innovation framework. These components are 1) the value co-creation 2) the gender component 3) the Responsible Research and Innovation (RRI) component. Co-creation is becoming an established innovation paradigm, in which the end-user is involved in the business model through participating in different phases, in order to shape the final product (or service) and increase its value; the gender component is a key point in the societal debate nowadays, and it is referred -to the different role men and women may have in the business structure and -to how a given function may affect differently those two gender classes; RRI is a new paradigm that relies on the Quadruple-Helix framework, in which not only the end users, but the whole community (education, policy makers, civil society, and industry) are involved in the creation process. The experimental analysis to model this aspect takes into account businesses from several geographic areas and industries, and relies on the comparison between neural networks and regression analysis.

Contents

Introduction

The aim of this thesis is to model the relationship between gender, value co-creation, and articulation of innovation. This will be done by comparing "traditional" methods, i.e., linear regression models, with Artificial Neural Networks. The use of neural networks is increasingly widespread: they are algorithms that take inspiration from the functioning of the human brain. Their application is suitable when the data processed by a user (or other stakeholder) is not subject to any assumptions about the relationship being studied, and for this reason they are defined as black-box.

Value co-creation is a business paradigm that is becoming more and more important in the field of marketing and innovation: it can be defined as a marketing logic that proposes the involvement of customers (and others stakeholders) in the creation of product and services, and this way it makes the business able to satisfy the needs of a heterogeneous group of customers. The analyses performed in this thesis are aimed at assessing whether there is a correlation between the opportunities to articulate the interest on innovation and the practice of value co-creation activities of businesses coming from different scenarios. The literature on these topic shows an increasing attention to co-creation activities, and a possible positive relationship between the concepts of "co-creation" and "innovation" (see Chapter 1). According to related literature, it is possible to quantify the interest on innovation and co-creation by examining corporate communication practices, and counting the occurrence of a well-defined set of regular expressions, that are associated to these two aspects. This approach will be outlined and used in what follows.

The discussion about gender is nowadays a key-point for policy-makers, civil society organizations, academic institutions, and practitioners: as we will see in Chapter 1, gender refers to a social construction and it is defined and assimilated through the interaction with other people. As a consequence, there is a high probability that people expect a woman to behave according to their "femininity", while men according to their "masculinity". In this framework, innovation and entrepreneurship represent a relevant case study: on one hand, there is evidence showing that male are more apt to apply an innovation that requires the application of a new technology in their businesses; on the other hand, there is an increasing awareness of the women's innovation potential, due to the fact that the values brought into a business by women may influence the level of innovation. There are also other ideas suggesting that women are more apt to understand the time-varying customers' needs, and that gender diversity on boards represents a valid tool to develop a broader range of ideas and to innovate the product or service. In a nutshell, there is an ongoing debate on the innovation potential of women, its difference with their male counterpart, and their assessment.

In the last decades, thanks to the rapid development of new technologies, innovation has assumed a prominent role in the entrepreneurial world. Since technology is available to almost everyone and innovation spreads quickly, the figure of the innovator is also widespread among a greater and greater number of individuals; therefore, while in the past the innovator was seen as a masculine role, recently studies are emerging on the female contribution to this dimension [27]. For this reason it is important within a business to understand who makes the choices regarding innovation, by analyzing the characteristics of the team members assigned to carry out this activity. In addition, it is important to take into account how those who make the choices communicate innovation-related news to potential customers and stakeholders. In particular, recent works [60] have considered the gender aspect with regard to the diversity of top management: according to [32] the diversification of the corporate board of directors due to the presence of women and men leads to heterogeneity with reference to different aspects e.g. experience, knowledge, perspectives, working style. This is the reason why many businesses have recently aimed to form a corporate board of directors with a greater number of women than in the past, especially with regard to businesses operating in a highly innovation-driven sector.

Given the importance of the topic, a relevant number of studies have been conducted with the aim to investigate whether there is a relationship between the gender of decision makers regarding innovation and the intensity of innovation itself: many proxy can be used to assess the intensity of innovation, and amongst these approaches some author have suggested to measure the occurrence of pre-defined regular expressions regarding innovation on the communication media used by the different businesses, with a particular attention to the web-site [55].

Recent approaches have been proposed by Bendell, Sullivan and Hanek

[16]: they examined the dry cleaning industry, and investigated the different approaches of men and women to choosing whether to invest in a new technology; as a result, they concluded that in their sample, gender affects this choice: at the same level of revenues, women have a lower tendency to invest in new technologies in order to implement innovation. Starting from this findings, the objective of this thesis is to examine whether there is a relationship between the composition of the corporate board of directors in terms of men and women, the co-creation activities, and the perception of (or alternatively, the attention to) the concept of innovation, taking into consideration a greater number of industries.

In this thesis, we want to use the specific co-creation and gender features of businesses to model the extent to which businesses are interested to the innovativeness of their products and services, that can be considered a good indicator of the attention that a business devotes to innovation practices. From a conceptual point of view, we first decide how to quantify the three aspects of "gender", "innovation", and "co-creation"; then, we quantify these three aspects for all businesses belonging to all sets of data taken into account; finally, we use "co-creation" and "gender" as predictors to build a model on "innovation".

As for the "gender" component, accordingly to the main literature we can easily quantify it by counting the number of male and female in the board of directors; as for the "interest on innovation" and "co-creation", it is possible to quantify their aspects by examining corporate communication practices, and counting the occurrence of a well-defined set of regular expressions. Among the indicators of results related to the innovation of businesses (i.e., the degree of articulation of innovative aspects), value co-creation practices have also been identified. For this reason, the investigation of the relationship between co-creation and interest on innovation does not need any initial assumptions about its specific functional form. We will start by quantifying the occurrence of these regular expressions on the web-sites of businesses from different sets of data, with the aim to detect similarities and differences amongst these sets of data. The sets of data taken into account are 5 sets of data composed of businesses belonging to stock indices (that represent an overall indicator a given country), and one sets of data is from a specific industry, i.e. Open-Source development focused businesses. Then, we will use linear regression and neural networks to model the relationship between gender, innovation and co-creation articulation.

In this thesis we aim:

- to investigate the relationship between co-creation and (articulation of) innovation over different benchmarks;
- to investigate the relationship between gender and innovation over different benchmarks;
- to build an approach to model or predict innovation on the basis of co-creation practices and/or gender aspects.

The thesis is structured as follows:

Chapter 1 will introduce the main concepts that will be subject of observation in the thesis: innovation and its representation in the businesses web activities; value co-creation, defined as a marketing paradigm in which the customer play an active role in the creation of value; and the gender component, which is more and more a key element of societal discussion for many different stakeholders. The related literature relevant for our analysis will be explored, along with the regular expressions that represent the building blocks of our approach.

Chapter 2 will describe the data used in our experimental phase: we will start by introducing the concept of web-scraping, and to detail the web-scraper that we have used for data collection. Then, we will describe the resulting five sets of data, that consist of real data coming from different countries.

Chapter 3 and Chapter 4 will give to the reader an introduction to the computational tools that will be used in the experimental phase: linear regression and Neural Networks.

Chapter 5 will deal with the analysis of our data through the application of linear regression models.

Chapter 6 will deal with the analysis of our data through the application of Artificial Neural Networks. Finally the results of linear regression models and neural networks will be compared.

Chapter 1

The background of our analysis: innovation, value co-creation and gender

In this chapter we are outlining the background of our analysis: we are discussing of innovation in Section 1.1, then value co-creation is described in Section 1.2, and finally we are dealing with the gender components in Section 1.3

1.1 Innovation

Innovation has been the focus of many debates over the years: it is becoming more and more interesting due to the progress of technology and society, as it plays a central role within the strategies adopted by businesses [22].

According to Aulet [9] the word *innovation* is the result of the product of two other terms, so we can think of this phenomenon as a mathematical product in which innovation is the product between invention and commercialization: Innovation = Invention $\times$ Commercialization.
From this product it is possible to observe that if one of the two elements (invention or commercialization) is equal to zero then also innovation does not exist: the presence of only one of the two is necessary but not sufficient.

Innovation has been defined in many different ways, so one of its limitations can be identified in the fact that there is no common definition for everyone according to which it is possible to identify its true nature [12]. Amongst the different definitions of innovations proposed by the literature, we find the following:

- According to Schumpeter [112], innovation is "the first introduction into the economic and social system of a new product, process or system";
- Thompson [119] defines innovation as "the generation, acceptance and implementation of new ideas, processes products or services";
- According to Van de Ven [121] innovation is "something that is perceived as new to the people involved, even though it may appear to others to be an imitation of something that exists elsewhere".

As mentioned earlier, innovation is a fundamental pillar of a business' strategy: the term "strategic innovation" was introduced in 1995 by Baden-Fuller [10] and was defined as a combination of actions aimed at innovating an organization, which could be promoted by both incumbents and new market entrants. This process served to improve the balance of certain trade-offs such as quality/productivity and efficiency/variety: thanks to innovation it is not necessary to completely sacrifice quality in exchange for productivity (and vice versa) and efficiency in exchange for variety (and vice versa).

It can be stated that innovation represents the heart of the competitive game, as it allows businesses to implement change and differentiate themselves from traditional competitors and the traditional way of doing business in a certain industry [22]. By taking advantage of the opportunities for innovation provided by changes in technology, markets, structures and dynamics, businesses are able to meet the changing needs of customers and to maintain their competitive position [12].

Moreover, according to Schumpeter [112], businesses need to innovate in order to renew and maintain the value of their assets: innovation is the beating heart of businesses and it is essential to ensure survival and growth in a competitive world [132]: therefore, innovation plays a central role in creating value and sustaining a competitive advantage. In addition, innovation is able to ensure the presence within an organization of a process of renewal and therefore growth [18] since this concept implies a process of change [33]. In particular, innovation can be divided into incremental change and discontinuous change according to its degree of intensity: in addition to the fact that it is not easy to find businesses that apply radical innovation and become market drivers [81], once applied it is essential to be able to manage both incremental and radical changes at the same time and therefore acquire the ability to be ambidextrous [100].

Different kinds of innovation can exist within a business' strategy. A first distinction was made by Drucker [45] in 1954 who identified two major types of innovation in relation to the change triggered in the business model, formed by the target (who), the offer (what) and the chain of processes (how):

- *product* or *service* innovation, which is related to "what" the business offers to customers;
- *process* innovation, which is related to "how" the business provides the product or service.

These same types of innovations have also been subsequently addressed by other authors [33] [51] who have referred to innovation in the area of new products, services, materials, resources, capabilities, and processes.
Later, *strategy innovation* was defined as a reconfiguration of the business model [85], a redefinition of a business' organization, a result of the ability to identify the right combinations of resources.

On the basis of the intensity of the innovation implemented, it is possible to identify other three kinds [22]:

- *sustaining innovation*, which refers to improvements in performance that are most relevant to existing customers;
- *low-end disruption*, which refers to strategies that aims at providing a low cost offering;
- *new market disruption*, which refers to the creation of a new market space.

Regarding *new market disruption*, the creation of a new market space has been associated by Kim and Mauborgne [71] to a value innovation, able to break the trade-offs between cost and differentiation. Subsequently, the same authors have addressed this issue by introducing the "blue ocean strategy" [72], associated with a radical innovation promoted by a business.

While presenting some distinctions, within the sphere of innovation it is possible to identify some recurring themes applied in different sectors [28]: these themes usually have a history and their trend can be predictable. One of the most recurrent themes is the transition from the simple offer of a product or service ready to the involvement of the customer in an experience; this experience can be of different types and among these it is possible to identify

the value co-creation, in which businesses involve their customers in order to create together a new product or service.
To this end, this thesis will analyze the possible relationship between value co-creation and the degree of innovation of businesses. In particular, we want to quantify the extent to which businesses are able to identify and communicate their new opportunities related to innovation. This is useful to understand businesses' perceptions of innovation with respect to products, services and processes.

Recently, there has been a significant growth in the amount of unstructured data in reference to digitalized texts. The elaboration of texts proves to be more and more complex and time-consuming, so the adoption of text mining techniques for the analysis of textual data has proved to be very useful in order to simplify and improve the activity of researchers [8] and other stakeholders.

Text mining refers to techniques capable of analyzing a unstructured text (e.g. a narrative text), processing the information contained within it and transforming it into structured content, i.e., following a certain scheme [65]. Through text mining, the data contained in a text is divided and classified according to the topics and subtopics to which it belongs, so as to have a clearer and more easily readable idea of the content of the document under consideration.
The text mining techniques may resort to Artificial Intelligence (AI) [43] and the application of these techniques has also been carried out in the field of research on innovation, leading to analyze the innovation according to different aspects [95]. It is possible to state that innovation can be studied from different points of view, such as product, service or process innovation [120]. Moreover, with the fast diffusion of new technologies, the development of innovations is even faster, so a totally manual analysis could be complex and not completely exhaustive.
In this thesis we aim at analyzing such different dimensions of innovation and whether they can be influenced by other variables such as the gender component and the co-creation activities. We will use a simple text mining procedure to assess the innovation dimension with regard to its web articulation. According to Cebi [26], businesses' online communication of innovation-related content influences customers' and other stakeholders' perceptions of innovation. Starting from this hypothesis, we want to investigate the amount of online content about innovation, which from now on we will call *articulation of innovation*. Articulation of innovation will be used as a metric to measure the degree of innovation of businesses in the analyses

that follow in this thesis. It will be measured through the calculation of the frequencies of some innovation-related keywords on the businesses' websites: related details are provided in Section 2.2.

1.2 Value co-creation

In this section we will address value co-creation: given the prominent importance of this issue, the aim of this section is to explain its main features. Subsequently, in this thesis co-creation will be treated as a variable in several analyses in order to explain its influence on business innovation.

As stated in the Introduction, value co-creation is becoming more and more important in the sphere of marketing and innovation in businesses: it can be defined as a marketing paradigm able to satisfy the needs of a more heterogeneous group of customers [124], through the involvement of customers and other stakeholders in the creation of the final product or service.

Traditional business paradigm relies on products already prepared and designed by the supplier and bought by customers; value co-creation allows the customer to be actively involved in the design and experience phase of the product or service [101]: therefore, co-creation promotes transformation of the customer from a *passive* user who uses the product or service to an *active* participant in the creation of its value [104].

Achieving value co-creation requires a good collaboration between the business and the customer and a key aspect of this paradigm is the relationship between these two entities: the aforementioned collaboration is a form of dialogue that should be interpreted as a process of learning together to achieve a shared goal for both parties [11].

For this reason, the involvement of customers in the design of the product or service should not be seen as a mere engineering process, but rather as an interactive process between businesses and customers, not necessarily planned in every aspect and often undertaken partly unconsciously by the customer [101]: when customers are partly unconsciously involved in value co-creation process it is more likely that they try to do their best as they think only of making the best desired product without thinking about business' interests [77]. Customers' preferences are fundamental elements which shape value co-creation process [123]: in order to satisfy these preferences customers need to receive from the business information, knowledge, skills and resources to be used during the process; in addition, the business must be able to influence value co-creation process in order to enable customers to make the most efficient use of the resources [98].

In this context it is possible to state that the advancement of technologies plays a very important role in terms of 1) customers being able to receive information 2) businesses providing customers with tools to participate in the value co-creation.

Regarding aspect 1), thanks to the advancement in Information and Communication Technology (ICT), more and more efficient technological tools allow individuals to receive news and information in real time and at a global level regarding any kind of thing. As a consequence, people become more aware of their needs to be satisfied by a product or a service [39].
With regard to aspect 2), it should be noted that the interaction between the business and the customer during the value co-creation process can be guaranteed by a channel of communication between them, and this channel often consists of an online platform [111]. The development of new technologies has improved the efficiency of these online platforms, making the collaboration between businesses and customers easier and faster.

In this thesis we adopt the definition of co-creation introduced by di Tollo, Tanev, Liotta and De March [40] [42] [41] [39] in which value co-creation has been identified as a single concept.

Please notice that in some articles [106] [50] the concept of co-creation has been divided into value co-creation and co-production, referring to the former for the involvement of the client in the phase of use and consumption of the product or service and to the latter for the involvement in the phase of design and production.

It is possible to identify some important aspects in value co-creation [101]:

1. The processes of value creation with the client, which can be distinguished if it is a business-to-customer (B2C) or business-to-business (B2B) relationship. In the first case, it is necessary to involve the activities of the individual in the creation of value, while in the second, the processes of another organization must be involved;

2. The processes of value creation with the supplier: this coincides with the concept described in the first point but seen from an opposite point of view;

3. The meeting processes between supplier and customer: in order to take advantage of collaboration opportunities for value creation, there needs to be a meeting point in which to establish relationships between those involved in the process; as explained earlier, these meeting points are often online platforms.

Aspect 3 states that value co-creation meeting points serve to leverage value co-creation opportunities. Choosing to take advantage of these opportunities is part of a business' strategic plan and depends on some factors, i.e., the industry in which the business operates and the customer segment served [101]. According to Payne, Storbacka and Frow [101] it is possible to identify three types of opportunities:

- Opportunities that have arisen thanks to new technological developments: with the advancement of technology comes new ways of collaboration between customers and suppliers, as well as an easier and faster exchange of information and resources;

- Opportunities born from changing industries: the transformation of industry boundaries brings opportunities to develop new skills and knowledge and innovate one' s organization [29];

- Opportunities arising from changes in customer preferences: as customers' lifestyles change, so do their preferences, and this represents an opportunity to create new value in the products and services offered.

Therefore, an entrepreneur who wants to practice value co-creation has to face the challenge of recognizing how to take advantage of the right entrepreneurial opportunities [114].

Our previous discussion is useful to understand that collaboration aimed at creating value is not a simple process. As already mentioned, however, the adoption of value co-creation shows a growing trend since its application leads to benefits for both the customer and the supplier.

With regards to customers, the value co-creation process is useful as they can get the product or service that actually meets their preferences [133]. Moreover, through collaboration they feel actively involved in the production process and this stimulates trust and loyalty towards the supplier [101].

From a supplier's perspective, value co-creation can help as it more clearly highlights customer preferences and provides the opportunity to create the products and services they want, thus meeting their interests [124]. In addition, collaboration with the customer can lead to the creation of new products and services that do not yet exist on the market: this leads the business to gain a competitive advantage and a greater degree of innovation [133]. In this regard, it has been stated by several authors [79] [78] [94] that the involvement of customers in value co-creation activities has a positive influence on the results of innovation; in particular, this activity is able to reduce innovation costs and time-to-market and to increase the quality of the new product or service and the business' development skills. For this reason, the development of new platforms that allow collaboration in the creation of value is considered increasingly important and placed among the key points of the strategic plan of businesses [103] [94].

As the increasing importance given to co-creation activities has been stated in the literature [124] and a possible positive relationship between the concepts of value co-creation and innovation has been found [39] [41] [42] [40], one of the main purposes of this thesis to investigate this relationship and contribute to the research.

1.3 Gender

While the sex of a person is a feature defined by biology, gender is something that refers to a social construction: it is a cultural aspect [83] and it is assimilated through the interaction with other people. Therefore, individuals learn how to behave according to their gender, and they also learn the attitudes to avoid in order not to be inappropriate [130].
As a consequence, there is a high probability that people expect a woman to behave according to their "femininity", while men according to their "masculinity" [128]. Therefore, there are stereotypes of the role attributed to both genders [48] and these can be classified as 1) *descriptive gender stereotypes*, that refer to what a man or woman is like, and 2) *prescriptive gender stereotypes*, that refer to the socially required behavior of woman and a man.

The elements just described can shape the behavior of individuals and have an influence on their decisions: men and women are partly influenced by what is *normal to do* for the society according to "masculinity" or "femininity". With this regard, a relevant example is given by the segregation of the labor market, i.e., the lack of males or females in a certain sector or profession [3].

The behavior of individuals is also affected by the context in which they find themselves because some contexts seem to require specific attitudes and roles. The reaction of people who find themselves in this circumstance is usually to adapt themselves to the context and behave as it requires, assuming traits that recall masculinity or femininity [24].

A relevant context where individuals' behavior is influenced by the context itself is the entrepreneurial world: from a psychological and competence point of view, it requires leadership skills [131] [92], which are identified in typical masculine traits such as aggressiveness, risk-taking and autonomy [5]. On the contrary, femininity seems to be attributed to a warm, calm and communal attitude and therefore, if a woman intends to undertake an entrepreneurial activity, she should assume those masculine traits just described [1]. Furthermore, according to the Global Report 2020/2021 [21] compiled by the Global Entrepreneurship Monitor (GEM), men involved in entrepreneurial activities are significantly more than women. In particular, the percentage of men entrepreneurs exceeds that of women in almost the entire world in the Total Early-stage Entrepreneurial Activity (TEA), which measures new entrepreneurial activities undertaken during the year; the exceptions are represented only by six countries (Kazakhstan, Indonesia, Oman, Saudi Arabia, Togo and Angola).

It should be noted that on top of external factors, such as context and expectations from society, there are other factors that influence and shape the choices made by men and women differently. In particular, a crucial feature is found in the degree of risk aversion which is manifested differently in the two genders [46] [127]. This aspect strongly affects the behavior of individuals and it is evident in the field of entrepreneurship where risk-taking choices such as evaluation and exploitation of different opportunities is required: women are more risk-averse and therefore tend to evaluate less positively such opportunities. Logically, this is a very relevant reason why there is less female entrepreneurship than male [73].

In the context of entrepreneurship there are significant differences between the two genders in innovation-related issues [97]. Innovation concept is characterized by 1) risk-taking component and by 2) implementation of new technologies.

Regarding the aspect 1), it has been just mentioned that women are more risk averse: in the case of having to implement an innovation, it is not possible to foresee perfectly what may happen, either positively or negatively, because it is something new; for this reason it is necessary to analyze the available opportunities and decide whether it is worth the risk to exploit them. Therefore, similar to what has been described for the exploitation of entrepreneurial opportunities, also in this case women are more risk averse.

The aspect 2) can be linked to the different study choices made according to gender. In fact, it has been shown by multiple studies [68] [23] [15] that in STEM (Science, Technology, Engineering, and Mathematics) education the presence of women significantly lower compared to men. As a result, more men acquire training and familiarity with technology and are therefore more likely to apply an innovation that requires the application of a new technology in their businesses.

Therefore, the sphere of innovation has historically been perceived as something that belongs to “masculinity": some works have shown a strong link between innovation and “masculinity" [126]. Women are often not seen as promoters of innovation activities [86], and there are several reasons for this. One possible reason is that women are tied to family commitments and responsibilities and therefore they have less time to develop innovative ideas. Another reason is related to the concept of context described earlier: since women are aware that they are not seen as promoters of innovation, their ideas for innovation tend to be inhibited [30].
Recently, theories according to which innovation is linked to masculinity have been partly re-evaluated: there is a growing focus on the role of women in innovation, particularly in businesses that are strongly innovation-driven [87]. Through the analysis of gender diversity in top management at businesses, some implications emerged [84]. Among these implications, a positive influence of the presence of women on the board of directors on innovation activities has been demonstrated [96]. Moreover, the presence of women on board of directors leads to greater heterogeneity in the group, in which there is consequently a greater diversity of perspectives, experiences, work styles, knowledge and skills [32]. With regard to skills and knowledge, it is found that women are more adept at recognizing customer behaviors and expectations; this leads to the identification of innovative products, services and processes that more closely reflect customer needs [69, 89].

For the reasons just described, it is important to take into account the composition of the corporate board of directors, considering the male and female components: it can affect the quality and quantity of innovation-related decisions taken by the group. Therefore, in this thesis we aim to investigate whether the composition of corporate board of directors influences the choice of innovation-related information communicated on the business website.

1.4 Related works

One of the principal purposes of this thesis is to analyze the relationship between co-creation and businesses' articulation of innovation: value co-creation is an emerging topic within marketing strategies [78] and it represents a growing trend also due to the advancement of technologies regarding participatory platforms where co-creation takes place. According to the main literature on the topic [78, 80, 93], the practice of value co-creation activities implemented by businesses should have a positive impact on the results concerning the sphere of innovation.

In what follows we outline the research ideas that will be useful to understand our contribution. This research ideas have as their main theme the research of the relationship between co-creation and innovation and they will be briefly described. In particular, we will present the origin and development of the metrics used to measure the two aspects (i.e., innovation and co-creation) and the approaches and tools used to calculate the results.

In order to get a measure of the two aspects, the frequency of keywords related to the concept of co-creation and innovation was searched with appropriate tools within the websites of a large sample of businesses. This approach was introduced by Ferrier [52] in 2001 who used the keywords searching method in order to analyze the sequence of competitive actions which describes a business' strategy. The results of this work showed that the keyword division of a topic (i.e., business' strategy) is useful to understand better the topic itself and its several aspects. Later, the same method was applied to innovation-related topics by Hicks, Libaers and Porter [58] in 2006: the aim of this paper is to analyze the business models of innovative businesses, which were found by searching innovative terms on their websites. In order to research the theme of innovation on websites, keywords were constructed and the number of keyword occurrences was normalized according to the size of the websites. This method has proven useful in order to analyze innovation in businesses of all sizes; thanks to this approach, the authors were able to classify businesses based on the degree and type of innovation communicated on their websites. So the keyword research approach has proven to be a good indicator of the importance given by businesses to a concept like innovation: for this reason it has been used in the works that are described below and in the analyses that will be presented in this thesis.

This approach was applied for the first time to analyze value co-creation by Allen, Tanev and Bailetti [6] in 2009: in this paper some keywords related to value co-creation were tested in order to classify the different co-creation activities practiced by a large sample of businesses. The co-creation keywords are listed in Table 1.1.

The use of keyword research on topics related to co-creation was further developed by Tanev et al. [117] in 2011: in this work for the first time keyword research was useful to search for topics related to co-creation and innovation on businesses' websites. The results of the research were used to conduct a linear regression to investigate the relationship between the two aspects. Through this analysis, initial results were found regarding the influence of the practice of co-creation activities on the degree of innovation communicated by businesses on their websites.

In 2010 di Tollo et al. [39] introduced an ANN approach instead of a linear regression in order to investigate the relationship between value co-creation and innovation among a large sample of businesses: this represents a great step forward since ANN approach is able to analyze different kinds of relationships between two variables and do not rely on prior assumptions about the form of relationship between them, while linear regression can capture only linear relationships. The hypothesis to be tested in this paper is that businesses with a higher degree of involvement in co-creation activity have a greater number of possibilities, occasions and contexts in which to apply innovation in their products, services and processes: in affermative case the theory of the literature would be confirmed [103] [50].

The initial empirical results suggested a positive association between co-creation activity and businesses' perceptions of innovation, but the results found were not yet able to explain a cause-and-effect relationship between the two variables.

Subsequently, in their 2012 paper di Tollo et al. [41] further investigated the possible relationship between the degree of co-creation activity and technology-driven businesses' perceptions of innovation. In this work, the authors started again from the hypothesis that greater involvement in co-creation brings the possibility for businesses to develop a better articulation of their innovation in products, services and processes.

Again, the frequency in the businesses' websites of keywords related to the two variables was calculated and the relationship between the two was analyzed through an ANN approach.
In this work for the first time the application of SOMs (Self-Organizing Maps) was introduced for the analysis of these two topics. Thanks to this method businesses have been classified according to their degree of involvement in co-creation and innovation activities.

Table 1.1: The co-creation keywords in Allen et al. 2009.

Co-creation keywords in Allen et al. 2009
$(customer \vee user) \wedge (learn \vee learning)$
$(customer \vee user) \wedge (communities \vee community \vee network \vee networking \vee forum)$
$(customer \vee user) \wedge (suggest \vee suggestion \vee input \vee request \vee demand)$
$(customer \vee user) \wedge (dialog \vee dialogue \vee communicate \vee communication \vee conversation \vee contact \vee feedback \vee call \vee interact \vee \text{``}information\,sharing\text{''} \vee engage)$
$internal \wedge (expertise \vee resource)$
$cost \wedge (reduce \vee reduction \vee saving)$
$customer \wedge (partnerships \vee interaction \vee relationship \vee participate \vee participation \vee activity \vee action)$
$(design \vee process) \wedge (flexibility \vee flexible \vee adaptable)$
$(customer \vee user) \wedge (cooperate \vee cooperation \vee collaboration \vee partnership)$
$(customer \vee user) \wedge (riskmanage \vee management \vee control \vee assess \vee reduce \vee reduction \vee potential \vee exposure)$
$trust \vee honesty \vee integrity$
$(customer \vee user) \wedge (options \vee choice \vee choose)$
$integrated \wedge online \wedge services$
$customization \vee customize \vee customized \vee personalize \vee$ individualize $\vee \text{``}add\,feature\text{''} \vee \text{``}added\,feature\text{''}$
$(product \vee process) \wedge (modularity \vee modular \vee module)$
$ecosystem \vee \text{``}value\,network\text{''} \vee \text{``}value\,constellation\text{''} \vee \text{``}multiple\,partners\text{''} \vee \text{``}external\,contributor\text{''} \vee \text{``}external\,source\text{''}$
$(customer \vee user) \wedge (disclose \vee inform \vee disseminate \vee reveal)$
$(customer \vee user) \wedge (produce \vee assemble \vee manufacture)$
$(customer \vee user) \wedge (IP \vee \text{``}intellectual\,property\text{''})$
$(customer \vee user) \wedge (test \vee trial \vee beta)$

The application of ANN and SOMs methods is important because both do not rely on prior assumptions about the form of relationship between the two variables. This is the reason why the two approaches were applied again by De March et al. [36] in 2012 in order to investigate the relationship between the same two variables.

The results of this empirical analysis show that the degree of co-creation could be a valid indicator of businesses' innovation-related outcomes, although again a causal relationship between the two variables cannot be inferred.

In 2014 another work was carried out by di Tollo et al. [42] with the same main purpose: to search for a relationship between the intensity of involvement in co-creation activities and the degree of innovation of the businesses taken into analysis.
Once again, the main research was based on the ANN approach: in this regard, the authors stated that this approach is very useful due to its ability to generalize since research on innovation is a relatively new field; ANN approach works without taking into account preliminary assumptions, which cannot be done with certainty given the novelty of the research. In addition, this approach was found to have a high degree of flexibility, adaptation, and to perform well in forecasting.

In this work, a quantitative analysis was carried out on the relationship between the two variables co-creation and innovation, taking into consideration a larger number of businesses. While in previous works only OSS businesses were analyzed, in this case the lists of businesses of the main stock exchange indexes were considered: those lists were selected since they best represent the economic and financial situation of a country. This choice was made in order to generalize the results obtained with OSS (Operations Support System) businesses to broader business categories.

Also in this case, an analysis of the public data available on the businesses' websites was conducted, by searching for the frequency of regular expressions related to co-creation and innovation: in this way, the degree with which businesses communicate information regarding the two variables to potential customers and other stakeholders was analyzed.
Subsequently, in addition to an ANN approach and a correlation analysis to study the possible relationship between the two variables, a Principal Components Analysis was also conducted in order to identify emerging groups of regular expressions.

The results of this work show once again that it is not possible to establish that there is a causal relationship, but that the practice of co-creation can be a good indicator of innovation-related outcomes achieved by businesses.

The last article in reference to these issues is the one written by di Tollo et al. [40] in 2015 . The main purpose of this paper was to find a relationship between the degree of involvement in co-creation activities by businesses, the degree of articulation of their service value attributes and their innovativeness.

This relationship was researched through the use of different methods: Principal Components Analysis was used in order to identify the components of co-creation activities; an ANN approach and correlation analysis were used to actually analyze the aforementioned researched relationship; a K-means cluster analysis[1] and Self-Organizing Maps (SOMs) approach allowed to classify the businesses according to their degree of involvement in different co-creation activities, articulation of their service value attributes and their innovativeness.

The results of this work show the presence of a statistically significant relationship between the degree of involvement in co-creation activities by businesses, the degree of articulation of their service value attributes and their innovativeness.

Through the study of these works just described which are mainly focused on the analysis of a possible relationship between the practice by businesses of value co-creation activities and the interest shown by businesses on the theme of innovation, it is possible to note that the first results on this topic show that there may be a positive relationship between the two variables. For this reason it is interesting to deepen this type of studies and this thesis aims to make a further contribution to this type of research.

[1]K-means cluster analysis is an algorithm which allows the user to classify a group of objects on the basis of their characteristics.

Chapter 2

Our Data

In this chapter we are describing the data collected and used during the experimental phase. First, we are describing the tool used for collecting data in Section 2.1. This tool has been used to analyse the content of the official websites of businesses belonging to different sets of data, and to count the occurrence of a pre-defined set of regular expressions that are related to innovation and co-creation practices: these sets of data will be described in Section 2.2, and some further details about correlation analysis and data reduction will be outlined in Sections 2.3 and 2.5.

2.1 Web scraping

Web scraping is used in order to search for information on external websites regarding a specific topic. Through this technology, it is possible to collect the necessary information on several websites, and to use it for several purposes. In this thesis we have used the *PHP Web Scraper for Regular Expressions* implemented by Flavio De Jesus Matias at the University of Luxemburg [35]. This tool has been tailored to the data collection requirement, and has been used to collect information regarding the aspects related to value co-creation and articulation of innovation: it allowed to calculate the frequency of a defined set of regular expressions denoting innovation and co-creation activities in the five sets of data that will be described in Section 2.2. The web scraping tool used consists of two parts:

- Server-side (back-end): it receives and processes the user's request and then provides a response. This represents the part that is not visible to the user;

- Client-side (front-end): this is the web interface to the user, used to interacts with the web scraping tool through sending requests to the server-side.

The server-side implements the algorithms to perform the following tasks:

- to transform the source code of a website into text;
- to translate the conditions imposed by the user into logical conditions and to apply them to the text of the website;
- to compute the occurrencies of the afore mentioned conditions, and to save the results in a CSV file.

In order to run the server it is necessary to use the program “XAMPP". The user has to provide as input:

- the URL where to search for information;
- the regular expressions to search for.

The input is entered through the web interface which is represented in Figure 2.1: it shows a screenshot of the web scraping tool while processing the first keyword related to value co-creation in the Eclipse dataset (see Section 2.2).

The first thing to enter into the web scraping tool is the URL where the user wants to search for information. The user can choose between two options to accomplish this step:

- “Insert Website URL": with this option it is possible to enter the URL related to one website only;
- “Upload CSV file": with this option it is possible to insert a CSV file containing a list of URLs related to the websites the user wants to analyze.

For our experiments, we have used the option “Upload CSV file".
Then, the user has to enter the condition that he/she wants to search the websites for; two options are possible to perform this task:

- To insert in each space reserved for “Conditions" a single word and choose one of the “Connectors" in order to link this word to the next and previous ones via the “Add condition" button. Possible “Connectors" are: *AND*, *OR*, *ANDNOT*, *ORNOT*;

Figure 2.1: Interface of the PHP Web Scraper for Regular Expressions.

- insert in the space reserved for “Conditions" a string of words connected by connectors: the string is written entirely in a single space of “Conditions" and each connector has to be put between two symbols “+".

For our experiments, we have used the second option.

Then the web scraping tool allows the user to select three different levels of the URL structure to be explored. Once all these information has been entered, clicking on the “Start" button the analysis of the web scraping tool begins. During the analysis in the “Results" box all the results related to the entered URLs and conditions are sent to the standard output.
At the end of the analysis it is possible to download a CSV file with the

results: this CSV file contains the following information related to the conditions and the URLs selected before:

- Hits on first page;
- Total number of hits;
- Total number of pages;
- The ratio between Hits on first page and Total number of pages.

For all URLs analysed by the WEB scraper, we have stored the ratio between Hits on first page and Total number of pages. This will represent the occurrences of the regular expressions in the experimental analysis.

2.2 Sets of data

In this section we are introducing data that will be the objects of the experimental analysis detailled in Chapters 5 and 6. In this thesis we want to use the specific co-creation and gender features of businesses to model the extent to which businesses are interested to the innovativeness of their products and services, that can be considered a good indicator of the attention that a business devotes to innovation practices. For this reason three aspects have been taken into consideration in reference to businesses: the *gender component*, the *value co-creation* and the *articulation of innovation*.

The *gender component* has been considered as the number of male and female in the board of directors. Alternatively, one may use the ratio of women in the board of directors of the businesses to the total number of members, which indicates the heterogeneity of the group. This component is included in some contribution aimed to test whether board heterogeneity can lead a business to an increase in interest in innovation, as reported by [34] [108].

The *value co-creation* refers to the degree of customer involvement by businesses in the creation of new products and services. As described in Section 1.2, this aspect has been computed by searching for co-creation-related keywords on businesses' websites; those keywords are listed in Table 2.1 .

Articulation of innovation refers to the way and extent to which businesses intend to articulate the innovative sphere regarding their products, services and processes. In particular, the frequency with which businesses

Table 2.1: The co-creation keywords used in this thesis.

Co-creation keyword	
C1	$customer \lor user - dialog \lor dialogue \lor conversation \lor feedback \lor call \lor interact \lor information \lor -exchange \lor information - sharing \lor information - access \lor engage$
C2	$customer \lor user \lor forum \lor connect \lor network \lor networking$
C3	$lease \lor rent \lor license \lor self - serve \lor self - service$
C4	$customer \lor user - cooperate \lor cooperation \lor collaboration \lor partnership$
C5	$customer \lor user - suggest \lor suggestion \lor input \lor request \lor demand$
C6	$internal - expertise \lor resource$
C7	$customer \lor user - risk - manage \lor management \lor control \lor assess \lor reduce \lor reduction \lor potential \lor exposure$
C8	$customer \lor user - IP \lor intellectual - property$
C9	$customer \lor user - learn \lor learning$
C10	$product \lor process \lor service - evolution \lor evolve$
C11	$customer \lor user - experience$
C12	$customer \lor user - test \lor trial \lor beta$
C13	$integrated - online - services$
C14	$simulation \lor simulate \lor model \lor modelling \lor virtual - world \lor reference - design \lor reference - flow \lor demo - application \lor toolkit \lor tutorial \lor sdk \lor software - development - kit$
C15	$product \lor process - modularity \lor modular \lor module$
C16	$customer \lor user - produce \lor assemble \lor manufacture$
C17	$customer \lor user - options \lor choice \lor choose$
C18	$design \lor process - flexibility \lor flexible \lor adaptable$
C19	$customer - partnership \lor interaction \lor relationship \lor participate \lor participation \lor activity \lor action$
C20	$cost - reduce \lor reduction \lor saving$
C21	$customer \lor user - survey \lor review \lor voting \lor vote \lor rate \lor rating$
C22	$trust \lor honesty \lor integrity \lor transparency$
C23	$customer \lor user - disclose \lor inform \lor disseminate \lor reveal$
C24	$customer \lor user - dashboard \lor statistics$

publish comments about innovation on their websites is examined in order to understand how interested businesses are in talking about their innovation and communicating it to their potential customers and other stakeholders: this public communication through websites has strong power as businesses are able to shape the perceptions of potential customers and other stakeholders regarding the innovation implemented [26]. We assess this concept via a regular expression introduced by Tanev et al. [118], that allows us to calculate the frequency of innovation-related comments via the following regular expression: $new \wedge (product \vee service \vee process \vee application \vee solution \vee feature \vee release \vee version \vee launch \vee introduction \vee introduce \vee new - product \vee new - service \vee new - process \vee new - solution \vee product - launch)$. As specified by di Tollo et al. [41], it should be noted that this metric does not actually measure how many innovative products, services and processes are available in a business' offering, but it rather reflects how much the business wants to communicate its idea of innovation by emphasizing its ability to differentiate itself.

All these metrics (gender components, co-creation regular expressions occurrence, and articulation of innovation) have been calculated with reference to the year 2021 on 5 datasets:

1. All businesses listed in Open Source (OS) list associated with the Eclipse OS Foundation and businesses listed in the two websites Open Source Expert[2] and the Canadian Companies Capabilities Directory of OS Companies[3]. This dataset will be referred to as *Eclipse* and contains a list of businesses known to have a high degree of innovation; it has already been used by [41] [42] in order to analyze the relation between co-creation and innovation. In this list there are 287 businesses.

2. All businesses listed on NASDAQ (National Association of Securities Dealers Automated Quotation), which is the U.S. electronic stock market. In this list there are 98 businesses.

3. All businesses listed on FTSE100 (Financial Times Stock Exchange), which is the British stock market. In this list there are 95 businesses.

4. All businesses listed on DAX30 (Frankfurt Stock Exchange) which is the German stock market. In this list there are 30 businesses.

[2] http://www.opensourceexperts.com, accessed on January 15^{th}, 2021.
[3] http://strategis.ic.gc.ca/epic/site/icttic.nsf/en/, accessed on January 15^{th}, 2021.

5. All businesses listed on the CAC40 Index (French stock market). In this list there are 40 businesses.

We want to point out that we had initially planned to take into account three more sets of data:

1. A list of the most relevant hotel brands in the world; this dataset will be referred to as *Hotel*. In this list there are 163 businesses.

2. A list of the most relevant online travel agencies in the world; this dataset will be referred to as *Agency*. In this list there are 89 businesses.

3. A list of the most relevant software houses which provide services to the hospitality industry; this dataset will be referred to as *Software*. In this list there are 54 businesses.

For these sets of data, we were unable to collect the gender composition of the board of directors on more than 40 percent of the total sample size, and the sample size was rather small: hence we are not using them in what follows.

Tables 2.2 – 2.4 reports the main statistics for data about the co-creation and innovation regular expressions and about the number of male and female on the board of directors. Data have been collected on March-April 2021.

Each dataset includes the 24 keywords related to value co-creation listed in the Table 2.1. For each keyword its frequency on the businesses' websites has been calculated: the main statistics of the frequencies are reported in the Tables 2.2 – 2.4. In all datasets there is the number of women and men in the board of directors. From these, the percentage of women in the board of directors has been calculated. Finally in each dataset there is the regular expression referring to the articulation of innovation (described before). Also for it, the frequency in the websites of each business was calculated and the main statistics of the frequencies are shown in the Tables 2.2 – 2.4.

Figure 2.2 shows the graphical relationships between innovation and the number of men and women on the board of directors relative to each dataset. On the y-axis there are the frequencies relative to the articulation of innovation; on the x-axis there is the number of men on the board of directors; the size of the points is equal to the number of women on the board of directors. At first glance, it is not possible to identify a regular relationship between the variables.
Figure 2.3 shows the graphical relationships between innovation and the ratio of women to total in the board of directors for each dataset. On the y-axis there are the frequencies relative to the articulation of innovation; on the x-axis there is the percentage of women on the board of directors. Also in

Table 2.2: Main statistics of the response and explanatory variables taken into account in the empirical analysis for Eclipse and NASDAQ datasets.

	Variable name	Average	STD	Min	Max
Eclipse Number of businesses = 287					
	$C1$	5.01	7.47	0	67.26
	$C2$	2.99	4.41	0	35
	$C3$	2.43	5.92	0	81.04
	$C4$	2.06	3.10	0	22
	$C5$	2.82	3.74	0	22
	$C6$	0.49	0.97	0	4.89
	$C7$	6.16	9.38	0	79.64
	$C8$	2.00	3.32	0	24.00
	$C9$	3.38	4.87	0	22
	$C10$	2.96	5.87	0	80.63
	$C11$	2.33	3.49	0	22
	$C12$	2.51	3.95	0	23.89
	$C13$	4.25	9.98	0	146.78
	$C14$	7.05	13.31	0	127
	$C15$	1.77	2.70	0	22
	$C16$	1.86	2.93	0	22
	$C17$	3.19	3.80	0	22
	$C18$	1.84	4.99	0	53
	$C19$	1.47	2.56	0	18
	$C20$	0.54	1.34	0	11.09
	$C21$	2.18	3.33	0	23.48
	$C22$	0.31	1.55	0	24
	$C23$	1.87	2.94	0	22
	$C24$	1.91	2.98	0	22
	I	0.17	0.32	0	2
	$G1$	3.22	3.27	0	17
	$G2$	6.28	4.72	0	31
NASDAQ Number of businesses = 98					
	$C1$	6.11	8.26	0	47.67
	$C2$	2.88	4.82	0	29.79
	$C3$	2.99	4.54	0	27.74
	$C4$	1.75	2.83	0	19.33
	$C5$	3.02	3.88	0	23.00
	$C6$	0.37	0.68	0	3.68
	$C7$	5.48	7.67	0	40.00
	$C8$	1.75	2.91	0	20.00
	$C9$	3.06	4.28	0	27.00
	$C10$	2.75	3.85	0	20.00
	$C11$	2.06	3.34	0	20.33
	$C12$	1.92	2.72	0	15.06
	$C13$	3.91	5.81	0	24.48
	$C14$	4.64	15.67	0	152.67
	$C15$	1.51	2.90	0	22.33
	$C16$	1.63	2.85	0	19.67
	$C17$	3.38	3.59	0	22.33
	$C18$	1.26	4.63	0	44.33
	$C19$	1.65	2.68	0	17.33
	$C20$	0.39	0.68	0	3.09
	$C21$	1.84	2.85	0	16.25
	$C22$	1.84	2.85	0	16.25
	$C23$	0.55	1.29	0	10.07
	$C24$	1.59	2.68	0	15.00
	I	0.10	0.18	0	1.00
	$G1$	7.57	2.01	1	14
	$G2$	3.04	1.35	0	8

Table 2.3: Main statistics of the response and explanatory variables taken into account in the empirical analysis for FTSE and DAX datasets.

	Variable name	Average	STD	Min	Max
FTSE Number of businesses = 95					
	$C1$	5.50	5.31	0	21.04
	$C2$	1.49	2.01	0	8.51
	$C3$	1.88	3.19	0	17.89
	$C4$	1.13	1.80	0	8.69
	$C5$	1.55	1.84	0	8.44
	$C6$	0.54	1.28	0	7.00
	$C7$	4.77	4.17	0	19.46
	$C8$	0.90	1.64	0	8.83
	$C9$	1.72	2.77	0	18.39
	$C10$	1.77	2.26	0	8.83
	$C11$	1.18	1.64	0	8.47
	$C12$	0.86	1.52	0	8.38
	$C13$	2.77	4.08	0	23.57
	$C14$	3.45	4.78	0	31.91
	$C15$	1.03	1.62	0	7.67
	$C16$	0.84	1.49	0	8.49
	$C17$	2.31	2.65	0	15.00
	$C18$	1.07	2.26	0	14.75
	$C19$	1.24	1.84	0	9.00
	$C20$	0.73	1.63	0	14.00
	$C21$	1.55	3.06	0	24.86
	$C22$	0.69	1.60	0	12.67
	$C23$	0.79	1.52	0	8.43
	$C24$	0.84	1.60	0	8.46
	I	0.08	0.17	0	1.04
	$G1$	7.07	1.92	4	15
	$G2$	3.81	1.41	0	7
DAX Number of businesses = 30					
	$C1$	6.00	4.92	0	24.45
	$C2$	1.74	1.62	0.10	7.35
	$C3$	3.08	3.37	0	15.15
	$C4$	1.76	1.56	0	5.99
	$C5$	1.51	1.31	0.02	5.38
	$C6$	0.60	0.69	0	2.37
	$C7$	8.05	5.60	0	22.26
	$C8$	1.07	1.11	0	4.50
	$C9$	1.79	2.28	0	9.07
	$C10$	3.46	2.53	0	9.33
	$C11$	1.43	1.40	0	5.04
	$C12$	1.13	1.19	0	4.25
	$C13$	4.43	4.08	0	16.32
	$C14$	6.38	6.68	0	22.21
	$C15$	1.98	1.74	0	5.87
	$C16$	1.06	1.11	0	4.17
	$C17$	3.00	2.43	0	12.69
	$C18$	1.26	1.38	0	5.12
	$C19$	1.37	1.36	0	5.06
	$C20$	0.42	0.38	0	1.39
	$C21$	2.47	2.02	0	9.10
	$C22$	0.53	0.87	0	3.30
	$C23$	1.04	1.06	0	4.29
	$C24$	1.17	1.25	0	4.57
	I	0.11	0.16	0	0.89
	$G1$	5.53	1.89	1	9
	$G2$	1.17	0.87	0	3

Table 2.4: Main statistics of the response and explanatory variables taken into account in the empirical analysis for CAC dataset.

	Variable name	avg	std. dev.	min	max
CAC Number of businesses = 40					
	$C1$	5.88	5.23	0	20.16
	$C2$	2.19	2.80	0	9.87
	$C3$	1.73	2.48	0	8.85
	$C4$	1.62	2.28	0	10.58
	$C5$	1.87	2.31	0	8.42
	$C6$	0.87	1.51	0	7.96
	$C7$	4.83	4.92	0	24.01
	$C8$	1.33	2.04	0	8.42
	$C9$	2.02	2.75	0	11.33
	$C10$	1.47	2.23	0	12.16
	$C11$	1.71	2.60	0	10.58
	$C12$	1.45	2.25	0	8.42
	$C13$	4.65	6.78	0	28.28
	$C14$	4.66	5.98	0	26.74
	$C15$	0.76	1.59	0	8.30
	$C16$	1.36	2.05	0	8.42
	$C17$	2.79	2.65	0	11.67
	$C18$	0.87	1.39	0	5.59
	$C19$	2.00	2.29	0	8.42
	$C20$	0.20	0.26	0	1.00
	$C21$	1.76	2.11	0	8.42
	$C22$	0.20	0.32	0	1.23
	$C23$	1.33	2.07	0	8.42
	$C24$	1.41	2.14	0	9.42
	I	0.11	0.23	0	1.00
	$G1$	7.95	2.22	3	13
	$G2$	5.33	2.47	0	10

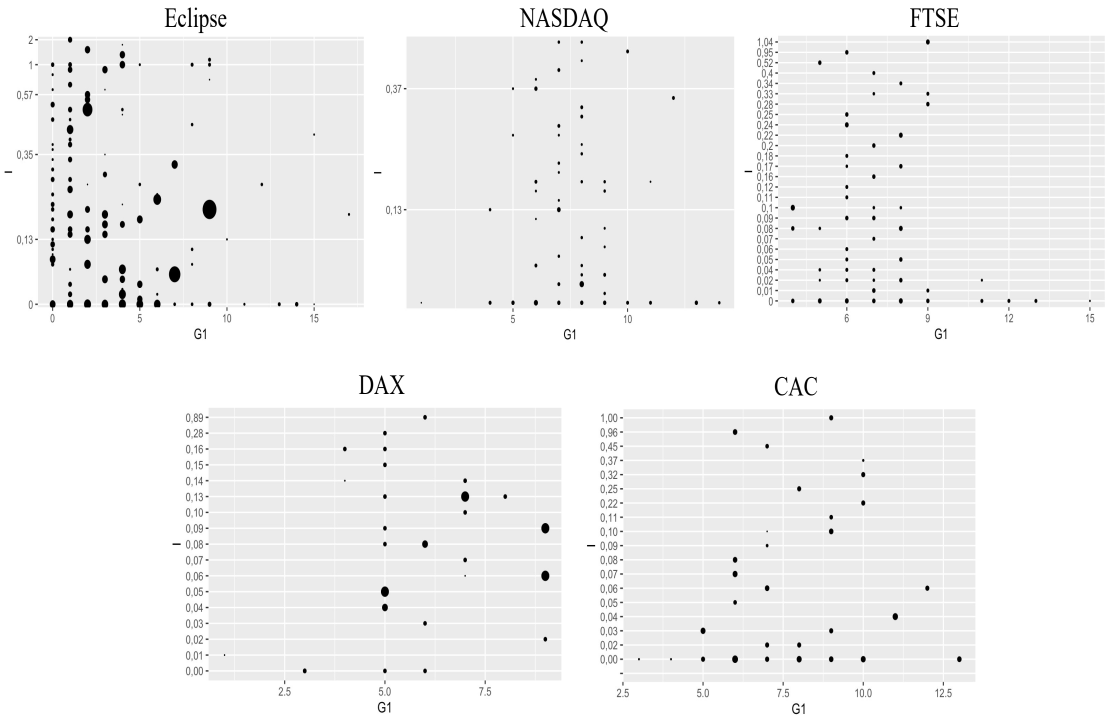

Figure 2.2: Scatter plot of the articulation of innovation vs number of men in the board of directors in the five sets of data taken into account. The point size is proportional to the number of women in the board of directors.

this case, at first glance, it is not possible to identify a regular relationship between the variables.

2.3 A first correlation analysis between gender and innovation

In this section we will present a correlation analysis between gender component and articulation of innovation in the eight datasets described in Section 2.2. We want to start by using, as a gender component, the ratio of women to total in the board of directors, since it has been identified as a potential indicator in the related literature. The purpose of this analysis is to understand whether the gender ratio on the board of directors and businesses' focus on

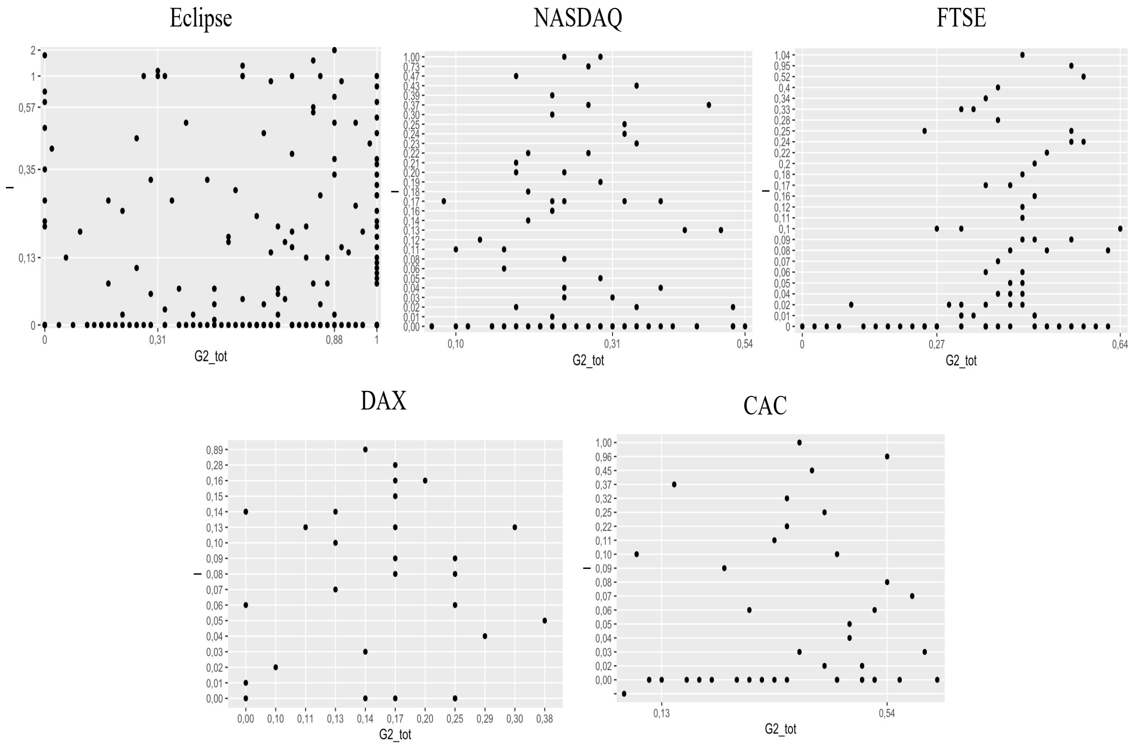

Figure 2.3: Scatter plot of the articulation of innovation vs ratio of women to total in the board of directors in the five sets of data taken into account.

innovation influence each other. Moreover, we want to investigate whether this influence is different in businesses that are known to be innovative and technology-driven (i.e., Eclipse) than in more traditional businesses. First of all, we had to compute the main statistics for the ration women to total, since it has not been computed in Tables 2.2 – 2.4: they are reported in Table 2.5.

It can be seen that the average ratio of women on the board of directors is highest in the Eclipse dataset. This seems to confirm what has been examined in the literature on gender and innovation, which states that in innovation-related businesses there are more women involved in the decision-making process than in the traditional businesses.

In addition, it is interesting to note that 71% of the businesses in the Eclipse dataset have an absolute majority of women on the board. This

Table 2.5: Main statistics of the ratio of women to total members of the board for the five datasets taken into account.

	Average	STD	Min	Max
Eclipse	0.66	0.32	0	1
NASDAQ	0.28	0.10	0	0.54
FTSE	0.35	0.11	0	0.64
DAX	0.16	0.10	0	0.38
CAC	0.38	0.14	0	0.63

figure marks a significant difference between the most innovative businesses and the traditional ones, counting that the second highest figure is 15% in the CAC dataset, while the other show small values, down to 0% in DAX.

We can also remark that the maximum female presence on the board is again found in the Eclipse dataset, where it is 100% (i.e., there are only women in the group). Also in the other datasets the maximum is referred to an absolute majority of women in the board, except for the list of businesses listed in DAX, where the maximum female presence is 38%.
Finally, with regard to the minimum values all datasets show 0, meaning no women on the board: the percentage of businesses without women in the board is between 1% and 2% (meaning really low) in Eclipse, NASDAQ, FTSE and CAC datasets.

These results seem to confirm that on a large scale entrepreneurship is still largely gendered, as has been stated by several authors [73] [90]. However, it should not be overlooked that it seems that compared to the past there is a greater presence of women in businesses where innovation is more intense: the same could happen for traditional businesses where there are some signs of change (such as the results referred to the absolute majority of women on the board).

Since in this thesis we are developing models to predict the articulation of innovation, we will draw some comments on this metric: by looking at Tables 2.2 – 2.4, we remark that the highest average is found in the Eclipse dataset: this confirms what was expected as these businesses are precisely those known to be more innovative. Moreover, the same can be seen for the maximum, in which the Eclipse dataset show the highest number. Instead for the minimum it is possible to observe that all datasets show the value zero: in every dataset there are therefore businesses that seem to have no interest in communicating their innovation to potential customers and stakeholders through their website. In particular, in all datasets the percentage of

Table 2.6: Correlation analysis between ratio of women to total in the board of directors and articulation of innovation.

	Pearson correlation		**Rank based correlation**	
	Pearson correlation coefficient	p-value	Correlation coefficient	p-value
Eclipse	0.02	0.76	0.10	0.13
NASDAQ	0.01	0.94	0.01	0.91
FTSE	0.23	0.03	0.31	0.01
DAX	-0.06	0.75	-0.10	0.59
CAC	0.09	0.60	0.07	0.66

Table 2.7: Correlation analysis between the number of men in the board of directors and the articulation of innovation.

	Pearson correlation		**Rank based correlation**	
	Pearson correlation coefficient	p-value	Correlation coefficient	p-value
Eclipse	-0.03	0.66	-0.08	0.22
NASDAQ	0.01	0.96	-0.02	0.86
FTSE	-0.03	0.75	-0.09	0.41
DAX	0.10	0.61	0.14	0.45
CAC	0.03	0.87	0.09	0.57

businesses that have zero frequency related to the articulation of innovation is about 50%, except for the 20% in DAX.

At this point, an analysis of the correlation between the ratio of women to total in the board of directors and articulation of innovation was conducted to see if they have an influence on each other: Pearson correlation analysis and a Rank based correlation analysis were done.

Table 2.6 shows that there is no correlation between the two variables for the all sets of data, since both Pearson correlation coefficient and Rank based correlation coefficient are very low and close to 0 (except for FTSE), while p-values are significantly greater than 0.05. Similar remarks can be drawn for the correlation analysis between: 1) the number of men in the board of directors and the articulation of innovation and 2) the number of women in the board of directors and the articulation of innovation. The results of these analyses are reported in Tables 2.7 – 2.8.

Table 2.8: Correlation analysis between the number of women in the board of directors and the articulation of innovation.

	Pearson correlation		**Rank based correlation**	
	Pearson correlation coefficient	p-value	Correlation coefficient	p-value
Eclipse	-0.03	0.63	-0.03	0.66
NASDAQ	-0.01	0.91	-0.04	0.67
FTSE	0.29	0.01	0.31	0.01
DAX	-0.03	0.89	0.01	0.96
CAC	0.07	0.67	0.06	0.72

2.4 Correlation matrices

Tables 2.9 – 2.13 show the Pearson Correlation Coefficients between all pairs of indicators taken into account for our experimental phase.

As far as the Eclipse Dataset is concerned, we can first identify some correlation coefficients greater than 0.90. They are present between the variables: C8 and C4; C8 and C5; C11 and C8; C16 and C4; C16 and C8; C23 and C4; C23 and C8; C23 and C16; C23 and C22; C24 and C4; C24 and C8; C24 and C16; C24 and C23. Therefore, a strong correlation is present among these variables. The lowest correlation coefficients, between 0 and 0.15, are observed between the variables related to the gender component (G1,G2,G2/TOT) and all other variables. Among these some coefficients have negative sign and most are found between G1 and all the other variables. Finally, a good correlation with coefficients between 0.60 and 0.90 is found among numerous co-creation variables.
Also in the NASDAQ dataset there are correlation coefficients greater than 0.90. They are present among the co-creation variables and in particular: C4 with C8, C11, C16, C19, C21, C22, C24; C8 with C11, C16, C21, C22, C24; C11 with C16, C21, C22, C24; C12 with C16, C24; C14 with C18; C16 with C19, C21, C22, C24; C21 with C24; C22 with C24. Correlation coefficients between 0 and 0.15 are present between variables G1,G2, G2/TOT and all other variables: most of these coefficients have a negative sign. Finally, many correlation coefficients between 0.60 and 0.90 are found among the co-creation variables.
In the FTSE dataset, correlation coefficients greater than 0.90 are found among the variables: C8 with C16, C23; C12 with C16, C23; C16 with C23, C24; C23 with C24. There are correlation coefficients between 0 and 0.15 between gender-related variables and all the other variables and between C6

and some other variables: some of these coefficients have negative sign. Regarding correlation coefficient between 0.60 and 0.90, the same remarks can be drawn as for the previous datasets.
As far as the DAX dataset is concerned, correlation coefficients greater than 0.90 can be found among the variables: C8 with C12, C16, C23; C12 with C16, C23; C16 with C23. Correlation coefficients between 0 and 0.15 are between G1,G2,G2/TOT and all other variables and between C3 and some other variables: most of them have negative sign. Again, the same remarks can be drawn as for the previous datasets for correlation coefficients between 0.60 and 0.90.
Finally, in the CAC dataset there are correlation coefficients greater than 0.90 between variables: C2 with C8, C9, C11, C12, C16, C17, C21, C23, C24; C4 with C8, C9, C11, C12, C16, C21, C23, C24; C8 with C9, C11, C12, C16, C17, C21, C23, C24; C9 with C11, C12, C16, C17, C21, C23, C24; C11 with C12, C16, C17, C21, C23, C24; C12 with C16, C21, C23, C24; C16 with C17, C21, C23, C24; C17 with C23, C24; C21 with C23, C24; C23 with C24; G2 with G2/TOT. Correlation coefficients between 0 and 0.15 can be found between gender related variables and all other variables: among them, there are a lot of coefficients with negative sign. Regarding coefficients between 0.60 and 0.90, the same remarks can be drawn as for the previous datasets.

From this initial correlation analysis it is not possible to make conclusions about the relationship between value co-creation, gender component and articulation of innovation. In this thesis this relationship will be investigated in more depth in order to understand its nature.

2.5 Hypothesis of Data Reduction

Some related literature [41, 40] introduces a Principal Component analysis over the co-creation variables in order to reduce the input size and to identify the number and set of keywords that appear together on the business' website. In these approaches, the resulting PCA components are ranked in terms of their loadings, and this allows us to set a correspondence between the components and some specific co-creation practices (and activities). In these related works, once the components have been identified, one has to select the number of components to use in the analysis. We have performed a preliminar PCA analysis over all sets of data at hand, that has shown the emergence of different components (and different percentages of the explained variance) over the different sets of data: this would lead to use different inputs for each sets of data under analysis. Furthermore, some communication

Table 2.9: Pearson Correlation Coefficient between pairs of indicators, set of data Eclipse.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	INN	G1	G2	G2/TOT
G2/TOT	0.05	0.08	0.05	0.09	0.07	0.13	0.07	0.11	0.01	0.01	0.10	0.10	0.10	0.01	-0.03	0.10	0.06	-0.03	0.04	0.01	0.09	0.01	0.10	0.10	0.02	-0.74	0.53	1
G2	0.08	0.09	0.03	0.06	0.05	0.09	0.08	0.08	-0.01	-0.04	0.06	0.05	0.07	-0.02	-0.11	0.08	0.06	-0.03	0.06	0.04	0.11	0.05	0.08	0.07	-0.03	-0.14	1	
G1	-0.07	-0.10	-0.06	-0.13	-0.13	-0.15	-0.08	-0.14	-0.04	-0.08	-0.15	-0.15	-0.11	-0.07	-0.07	-0.14	-0.11	-0.03	-0.08	0.02	-0.10	0.00	-0.15	-0.14	-0.03	1		
INN	0.46	0.49	0.33	0.44	0.42	0.23	0.44	0.46	0.43	0.45	0.45	0.49	0.36	0.34	0.41	0.42	0.37	0.23	0.25	0.29	0.36	0.04	0.41	0.42	1			
C24	0.69	0.76	0.34	0.96	0.87	0.25	0.62	0.93	0.77	0.56	0.88	0.87	0.57	0.49	0.49	0.98	0.86	0.31	0.87	0.28	0.89	0.19	0.99	1				
C23	0.70	0.75	0.32	0.98	0.88	0.22	0.61	0.94	0.77	0.52	0.89	0.88	0.54	0.47	0.45	0.99	0.87	0.29	0.89	0.29	0.90	0.19	1					
C22	0.23	0.26	0.28	0.22	0.19	0.19	0.29	0.22	0.23	0.18	0.19	0.19	0.29	0.65	0.24	0.20	0.21	0.74	0.26	0.08	0.21	1						
C21	0.69	0.69	0.28	0.88	0.80	0.21	0.57	0.85	0.71	0.49	0.80	0.79	0.51	0.48	0.44	0.90	0.80	0.31	0.84	0.28	1							
C20	0.56	0.34	0.37	0.35	0.26	0.39	0.65	0.28	0.41	0.54	0.30	0.41	0.46	0.37	0.38	0.28	0.34	0.27	0.18	1								
C19	0.61	0.72	0.33	0.86	0.79	0.12	0.48	0.82	0.71	0.46	0.78	0.74	0.49	0.45	0.46	0.89	0.82	0.34	1									
C18	0.41	0.41	0.34	0.33	0.27	0.25	0.45	0.33	0.31	0.47	0.30	0.30	0.41	0.86	0.61	0.31	0.32	1										
C17	0.67	0.71	0.37	0.85	0.81	0.22	0.59	0.82	0.76	0.53	0.78	0.77	0.52	0.46	0.48	0.87	1											
C16	0.71	0.76	0.33	0.98	0.89	0.22	0.61	0.94	0.78	0.52	0.89	0.89	0.54	0.48	0.47	1												
C15	0.57	0.55	0.41	0.48	0.47	0.36	0.58	0.47	0.45	0.84	0.49	0.47	0.59	0.67	1													
C14	0.57	0.58	0.46	0.51	0.44	0.37	0.63	0.47	0.46	0.59	0.43	0.51	0.52	1														
C13	0.68	0.57	0.40	0.58	0.54	0.45	0.69	0.51	0.53	0.69	0.53	0.53	1															
C12	0.71	0.69	0.36	0.89	0.80	0.24	0.66	0.83	0.74	0.53	0.80	1																
C11	0.73	0.78	0.32	0.88	0.89	0.21	0.66	0.95	0.79	0.59	1																	
C10	0.73	0.67	0.60	0.56	0.57	0.45	0.75	0.57	0.56	1																		
C9	0.70	0.71	0.46	0.78	0.79	0.23	0.63	0.81	1																			
C8	0.75	0.80	0.35	0.92	0.91	0.21	0.66	1																				
C7	0.84	0.69	0.51	0.68	0.67	0.47	1																					
C6	0.42	0.29	0.27	0.27	0.24	1																						
C5	0.75	0.77	0.33	0.89	1																							
C4	0.74	0.77	0.35	1																								
C3	0.50	0.51	1																									
C2	0.77	1																										
C1	1																											

Table 2.10: Pearson Correlation Coefficient between pairs of indicators, set of data NASDAQ.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	INN	G1	G2	G2/TOT
G2/TOT	-0.15	-0.06	-0.06	-0.05	-0.22	0.09	-0.08	-0.07	0.05	0.03	-0.08	-0.04	-0.12	0.03	0.06	-0.05	-0.07	0.03	-0.07	-0.09	-0.08	-0.08	-0.08	-0.09	0.01	-0.32	0.86	1
G2	-0.16	-0.11	-0.04	-0.12	-0.27	0.07	-0.13	-0.15	-0.02	-0.02	-0.14	-0.09	-0.12	-0.06	0.00	-0.12	-0.14	-0.05	-0.16	-0.15	-0.14	-0.14	-0.10	-0.16	-0.01	0.09	1	
G1	0.07	-0.06	0.06	-0.10	-0.21	0.02	-0.03	-0.10	-0.09	-0.06	-0.05	-0.06	0.08	-0.12	-0.07	-0.10	-0.05	-0.13	-0.13	-0.03	- 0.07	-0.07	-0.01	-0.09	0	1		
INN	0.32	0.53	0.38	0.39	0.28	0.25	0.41	0.40	0.39	0.37	0.37	0.39	0.45	0.17	0.26	0.38	0.37	0.17	0.37	0.24	0.41	0.41	0.12	0.43	1			
C24	0.79	0.88	0.46	0.95	0.82	0.23	0.84	0.96	0.83	0.78	0.93	0.90	0.73	0.61	0.65	0.97	0.84	0.60	0.88	0.44	0.96	0.96	0.37	1				
C23	0.49	0.37	0.13	0.41	0.39	0.15	0.36	0.41	0.38	0.30	0.39	0.36	0.22	0.45	0.40	0.39	0.45	0.41	0.35	0.16	0.36	0.36	1					
C22	0.79	0.86	0.46	0.95	0.85	0.29	0.88	0.96	0.84	0.82	0.95	0.90	0.72	0.63	0.72	0.96	0.84	0.64	0.86	0.47	1	1						
C21	0.79	0.86	0.46	0.95	0.85	0.29	0.88	0.96	0.84	0.82	0.95	0.90	0.72	0.63	0.72	0.96	0.84	0.64	0.86	0.47	1							
C20	0.37	0.35	0.24	0.44	0.38	0.45	0.60	0.44	0.51	0.41	0.43	0.49	0.50	0.17	0.24	0.43	0.48	0.14	0.44	1								
C19	0.76	0.83	0.46	0.91	0.81	0.20	0.79	0.90	0.80	0.73	0.85	0.82	0.74	0.68	0.66	0.91	0.80	0.65	1									
C18	0.61	0.67	0.41	0.73	0.61	0.15	0.57	0.72	0.69	0.62	0.68	0.53	0.43	0.97	0.86	0.73	0.64	1										
C17	0.75	0.80	0.52	0.87	0.77	0.33	0.80	0.86	0.82	0.75	0.85	0.81	0.68	0.65	0.70	0.86	1											
C16	0.80	0.89	0.49	0.98	0.84	0.22	0.85	0.98	0.87	0.80	0.94	0.90	0.71	0.74	0.74	1												
C15	0.67	0.66	0.58	0.77	0.64	0.37	0.64	0.73	0.73	0.87	0.72	0.60	0.58	0.82	1													
C14	0.61	0.69	0.41	0.73	0.60	0.10	0.60	0.73	0.70	0.59	0.67	0.53	0.44	1														
C13	0.70	0.70	0.49	0.71	0.61	0.39	0.72	0.72	0.69	0.75	0.72	0.67	1															
C12	0.75	0.80	0.41	0.89	0.76	0.28	0.79	0.88	0.86	0.73	0.86	1																
C11	0.78	0.84	0.44	0.94	0.81	0.28	0.85	0.93	0.84	0.79	1																	
C10	0.71	0.70	0.63	0.82	0.68	0.44	0.71	0.79	0.77	1																		
C9	0.71	0.77	0.45	0.86	0.71	0.39	0.80	0.87	1																			
C8	0.82	0.87	0.48	0.97	0.84	0.23	0.85	1																				
C7	0.71	0.80	0.41	0.85	0.75	0.41	1																					
C6	0.24	0.27	0.38	0.29	0.22	1																						
C5	0.76	0.75	0.42	0.85	1																							
C4	0.81	0.90	0.53	1																								
C3	0.42	0.50	1																									
C2	0.73	1																										
C1	1																											

Table 2.11: Pearson Correlation Coefficient between pairs of indicators, set of data FTSE.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	INN	G1	G2	G2/TOT
G2/TOT	0.21	0.13	0.16	0.24	0.05	0.20	0.34	0.17	0.08	0.17	0.19	0.18	0.20	0.16	0.13	0.20	0.20	0.01	0.15	0.01	0.06	0.09	0.17	0.12	0.23	-0.51	0.89	1
G2	0.25	0.14	0.15	0.17	0.08	0.29	0.25	0.09	0.04	0.17	0.14	0.09	0.23	0.10	0.16	0.14	0.14	-0.06	0.13	0.10	0.02	0.16	0.10	0.08	0.29	-0.15	1	
G1	0.01	0.05	-0.04	-0.19	0.07	0	-0.19	-0.16	-0.02	-0.01	-0.09	-0.17	-0.04	-0.13	0	-0.15	-0.18	-0.10	-0.04	0.24	-0.06	0.05	-0.14	-0.07	-0.03	1		
INN	0.39	0.36	0.22	0.27	0.21	0.34	0.35	0.29	0.20	0.28	0.27	0.29	0.26	0.27	0.20	0.30	0.25	0.07	0.26	0.08	0.13	0.25	0.30	0.25	1			
C24	0.44	0.81	0.26	0.80	0.69	-0.07	0.55	0.86	0.52	0.22	0.81	0.87	0.12	0.10	0.11	0.92	0.64	0.09	0.61	0	0.49	0.02	0.92	1				
C23	0.42	0.76	0.12	0.86	0.76	-0.04	0.57	0.93	0.56	0.28	0.89	0.95	0.14	0.08	0.17	0.96	0.70	0.10	0.63	0.01	0.49	0.03	1					
C22	0.30	0.11	0.14	0.11	0.06	0.30	0.31	0.02	0.16	0.56	0.12	0.03	0.63	0.42	0.71	0.08	0.12	0.39	0.21	0.08	-0.02	1						
C21	0.23	0.39	0.06	0.40	0.37	-0.07	0.33	0.45	0.29	0.17	0.44	0.46	0.14	0.08	0.11	0.47	0.41	0.08	0.33	0.09	1							
C20	0.21	0.26	0.23	0.13	0.35	0.06	0.24	0.02	0.09	0.38	0.05	0.01	0.24	0.15	0.20	0.01	0.25	0.10	0.10	1								
C19	0.50	0.64	0.23	0.66	0.56	0.14	0.63	0.63	0.41	0.31	0.71	0.65	0.15	0.27	0.28	0.65	0.54	0.19	1									
C18	0.30	0.24	0.07	0.21	0.23	0.14	0.37	0.07	0.51	0.51	0.19	0.07	0.34	0.82	0.55	0.12	0.29	1										
C17	0.48	0.63	0.28	0.81	0.67	0.06	0.70	0.67	0.58	0.47	0.75	0.68	0.48	0.23	0.36	0.71	1											
C16	0.44	0.77	0.21	0.84	0.76	-0.03	0.59	0.90	0.57	0.28	0.88	0.92	0.14	0.13	0.19	1												
C15	0.45	0.22	0.15	0.23	0.21	0.42	0.36	0.12	0.26	0.82	0.24	0.16	0.55	0.42	1													
C14	0.44	0.28	0.17	0.21	0.20	0.20	0.42	0.09	0.49	0.42	0.19	0.13	0.29	1														
C13	0.29	0.22	0.23	0.34	0.20	0.28	0.47	0.12	0.33	0.55	0.19	0.19	1															
C12	0.40	0.78	0.22	0.83	0.73	-0.02	0.57	0.89	0.57	0.26	0.86	1																
C11	0.45	0.77	0.19	0.84	0.76	-0.02	0.67	0.85	0.56	0.32	1																	
C10	0.47	0.37	0.37	0.36	0.33	0.29	0.46	0.26	0.31	1																		
C9	0.53	0.55	0.12	0.60	0.55	0.12	0.50	0.52	1																			
C8	0.43	0.78	0.17	0.82	0.71	-0.04	0.59	1																				
C7	0.54	0.59	0.27	0.70	0.53	0.26	1																					
C6	0.54	0.07	0.08	0.03	-0.06	1																						
C5	0.35	0.74	0.19	0.68	1																							
C4	0.49	0.68	0.23	1																								
C3	0.20	0.44	1																									
C2	0.48	1																										
C1	1																											

Table 2.12: Pearson Correlation Coefficient between pairs of indicators, set of data DAX.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	INN	G1	G2	G2/TOT
G2/TOT	0.01	-0.27	-0.19	-0.09	-0.18	0.22	-0.13	-0.21	-0.11	0.18	-0.10	-0.26	-0.13	-0.02	-0.10	-0.21	-0.38	0.04	-0.13	-0.20	-0.23	0.21	-0.21	-0.17	-0.06	0.09	0.85	1
G2	0.01	-0.20	-0.14	-0.08	-0.16	0.26	-0.07	-0.11	-0.11	0.15	-0.08	-0.15	-0.16	0.02	-0.13	-0.14	-0.23	-0.02	-0.16	-0.20	-0.19	0.21	-0.12	-0.14	-0.03	0.47	1	
G1	0.08	0.11	-0.43	0.01	-0.01	-0.01	0.13	0.10	-0.01	-0.10	0.03	0.02	-0.07	0.12	-0.10	0.04	-0.10	-0.10	-0.07	-0.03	0.09	0.03	0.06	-0.01	0.10	1		
INN	0.25	0.63	-0.11	0.48	0.40	0.36	0.51	0.28	0.60	0.31	0.47	0.33	0.36	0.59	0.49	0.34	0.23	0.62	0.44	0.47	0.16	0.11	0.29	0.23	1			
C24	0.67	0.73	0.15	0.55	0.63	0.12	0.20	0.76	0.45	0.39	0.78	0.74	0.59	0.17	0.33	0.76	0.50	0.50	0.45	0.19	0.18	0.33	0.77	1				
C23	0.28	0.78	0.06	0.70	0.88	0.20	0.18	0.99	0.63	0.22	0.89	0.97	0.26	0.31	0.34	0.99	0.56	0.54	0.43	0.32	0.34	0.08	1					
C22	0.71	0.08	0.07	0.13	0.22	0.18	0.45	0.08	0.10	0.55	0.18	0.07	0.33	0.24	0.49	0.06	0.02	0.23	0.29	-0.03	0.18	1						
C21	0.22	0.22	-0.11	0.39	0.34	0.11	0.34	0.33	0.28	0.27	0.39	0.36	0.29	0.44	0.41	0.35	0.28	0.35	0.33	0.28	1							
C20	0.27	0.38	-0.03	0.41	0.33	0.51	0.46	0.32	0.40	0.37	0.28	0.42	0.36	0.60	0.64	0.38	0.40	0.63	0.44	1								
C19	0.48	0.45	-0.01	0.55	0.54	0.29	0.73	0.41	0.69	0.58	0.53	0.47	0.38	0.49	0.83	0.44	0.56	0.60	1									
C18	0.38	0.64	0.05	0.61	0.56	0.46	0.39	0.52	0.57	0.50	0.75	0.61	0.62	0.57	0.67	0.58	0.53	1										
C17	0.19	0.62	0.57	0.40	0.42	0.34	0.41	0.55	0.32	0.47	0.55	0.66	0.52	0.24	0.49	0.56	1											
C16	0.29	0.80	0.05	0.70	0.88	0.22	0.20	0.98	0.65	0.22	0.89	0.97	0.30	0.36	0.36	1												
C15	0.49	0.39	0.03	0.57	0.47	0.54	0.73	0.32	0.58	0.70	0.42	0.44	0.41	0.72	1													
C14	0.31	0.33	-0.06	0.52	0.36	0.43	0.48	0.29	0.51	0.46	0.36	0.37	0.29	1														
C13	0.62	0.50	0.31	0.20	0.19	0.22	0.41	0.25	0.12	0.58	0.46	0.35	1															
C12	0.26	0.81	0.15	0.68	0.84	0.32	0.23	0.96	0.63	0.32	0.86	1																
C11	0.38	0.83	-0.02	0.73	0.82	0.16	0.27	0.88	0.64	0.29	1																	
C10	0.53	0.30	0.35	0.21	0.23	0.46	0.59	0.21	0.30	1																		
C9	0.32	0.59	-0.09	0.64	0.70	0.33	0.35	0.62	1																			
C8	0.28	0.78	0.06	0.68	0.87	0.19	0.19	1																				
C7	0.47	0.31	-0.05	0.33	0.30	0.24	1																					
C6	0.09	0.26	0.13	0.34	0.13	1																						
C5	0.34	0.71	-0.03	0.65	1																							
C4	0.27	0.56	-0.10	1																								
C3	0.03	0.09	1																									
C2	0.35	1																										
C1	1																											

Table 2.13: Pearson Correlation Coefficient between pairs of indicators, set of data CAC.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	INN	G1	G2	G2/TOT
G2/TOT	-0.01	-0.07	-0.03	0	0.02	0.06	-0.10	-0.04	0.01	0.03	0	-0.08	-0.04	0.15	0.03	-0.03	0.03	-0.07	-0.05	0.01	-0.02	0.17	-0.04	-0.01	0.09	-0.24	0.90	1
G2	-0.08	-0.04	0.03	0.03	0	-0.04	-0.08	-0.01	0.03	0.02	0.03	-0.03	-0.05	0.06	0.01	0	0.03	-0.12	-0.03	0.01	0.04	0.18	0	0.02	0.07	0.14	1	
G1	-0.09	0.16	0.13	0.14	0.03	-0.12	0.10	0.16	0.16	0.07	0.16	0.21	0.11	-0.06	0.03	0.16	0.07	0.05	0.11	0.09	0.19	0.09	0.16	0.17	0.03	1		
INN	0.20	0.42	0.09	0.34	0.20	0.09	0.36	0.27	0.29	0.34	0.39	0.28	0.35	0.33	0.14	0.27	0.28	0.39	0.25	0.08	0.29	-0.11	0.28	0.25	1			
C24	0.60	0.91	0.36	0.97	0.87	0.25	0.83	0.99	0.94	0.60	0.97	0.96	0.77	0.47	0.44	0.98	0.91	0.53	0.76	0.35	0.92	0.33	0.99	1				
C23	0.61	0.93	0.36	0.97	0.89	0.27	0.85	1	0.96	0.63	0.97	0.97	0.78	0.49	0.48	0.99	0.92	0.57	0.77	0.39	0.94	0.32	1					
C22	0.43	0.25	0.19	0.35	0.29	0.29	0.18	0.34	0.30	0.01	0.31	0.29	0.31	0.40	-0.08	0.34	0.43	0.25	0.41	0.12	0.42	1						
C21	0.72	0.90	0.37	0.93	0.83	0.33	0.85	0.94	0.92	0.60	0.91	0.92	0.79	0.58	0.45	0.94	0.88	0.55	0.79	0.50	1							
C20	0.51	0.47	0.09	0.37	0.41	0.41	0.50	0.38	0.55	0.47	0.35	0.51	0.54	0.48	0.55	0.41	0.45	0.53	0.45	1								
C19	0.75	0.81	0.38	0.76	0.74	0.55	0.77	0.78	0.77	0.54	0.78	0.76	0.76	0.63	0.35	0.78	0.82	0.56	1									
C18	0.62	0.64	0.18	0.55	0.58	0.36	0.57	0.56	0.59	0.63	0.59	0.61	0.62	0.77	0.56	0.57	0.54	1										
C17	0.68	0.89	0.36	0.90	0.80	0.50	0.81	0.92	0.92	0.51	0.91	0.88	0.76	0.58	0.34	0.92	1											
C16	0.62	0.93	0.36	0.97	0.89	0.30	0.85	0.99	0.95	0.63	0.97	0.97	0.78	0.51	0.46	1												
C15	0.39	0.52	0.14	0.44	0.56	0.19	0.62	0.45	0.48	0.90	0.40	0.46	0.39	0.46	1													
C14	0.85	0.58	0.19	0.51	0.54	0.71	0.57	0.50	0.53	0.56	0.52	0.48	0.54	1														
C13	0.63	0.83	0.38	0.79	0.69	0.33	0.78	0.78	0.82	0.58	0.80	0.81	1															
C12	0.58	0.93	0.31	0.94	0.87	0.24	0.83	0.97	0.95	0.59	0.94	1																
C11	0.59	0.93	0.39	0.95	0.87	0.27	0.84	0.97	0.94	0.61	1																	
C10	0.52	0.69	0.29	0.59	0.65	0.24	0.78	0.61	0.61	1																		
C9	0.61	0.92	0.35	0.94	0.84	0.38	0.83	0.95	1																			
C8	0.62	0.93	0.36	0.97	0.89	0.27	0.85	1																				
C7	0.70	0.86	0.29	0.80	0.77	0.35	1																					
C6	0.60	0.36	0.02	0.27	0.32	1																						
C5	0.59	0.85	0.37	0.86	1																							
C4	0.62	0.90	0.39	1																								
C3	0.22	0.39	1																									
C2	0.63	1																										
C1	1																											

models used by practicioners (i.e., the SISEM model, see Appendix 6.6) determines a relationship between each co-creation keyword and a particular personal attitude about co-creation, hence all keywords have a peculiar interest for using these model[4]. Since we want to devise a procedure that is general and robust with respect to different sets of data, that uses all data available and that explains the contribution of all indicators (through the parameter estimates) we have decided to avoid the PCA analysis to reduce the space of inputs. See Appendix 6.6 for an outline of the PCA devised in our preliminar experiments.

[4]The SISEM model is the model used by Thierry Delperdange Coaching (BE-LU), that hosted me for an internship in March-April 2021. Details about their use of the SISEM model can be found in the book Eric Montier: *Connaître les Moteurs qui vous propulsent dans la vie*, Image Publique Editions, 2017.

Chapter 3

Linear regression

In this chapter we introduce linear regression: it will be used in Chapter 5 to investigate the influence of the aspects of value co-creation and gender components on the articulation of innovation made by businesses. Thanks to this technique it will be possible to collect the first results on the relationship between the three aspects (i.e., value co-creation, gender components and interest of innovation) and compare them with those obtained by other approaches (i.e., Artificial Neural Network).

Linear regression is a statistical technique useful for evaluating and describing the relationships that exist between several variables under consideration. Thanks to its versatility, this method can be applied in many different fields (e.g. Biology, Physics, Engineering, Economics, Medicine etc.): it is considered a basic tool in data science and is widely used to estimate the expected value of a dependent variable Y calculated from the value of other independent variables $X_1, X_2, ..., X_k$.
The first applications of a primitive form of linear regression were used, discussed, and published by Adrien-Marie Legendre in 1805 [82] and Carl Friedrich Gauss in 1809 [54]. Both of them used the method of Ordinary Least Squares (OLS)[5] by applying it to an astronomical problem with the objective of determining the orbits of celestial bodies around the sun.

Linear Regression can be distinguished between *Simple Linear Regression* and *Multiple Linear Regression.* Section 3.1 will explain the main concepts behind the Simple Linear Regression, and Section 3.2 will introduce the Multiple Linear Regression.

[5]OLS is an optimization technique in which the user searches for a function which minimizes the sum of the squares of the distances between the observed data and the curve representing the function.

3.1 Simple Linear Regression

Simple linear regression studies the dependence on average that exists between two phenomena, looking for a function that can express that dependence.
The adjective simple indicates the fact that the dependent variable y (also called response variable) depends only on an independent variable x (also called predictor or regressor variable).
The regression is linear because the average dependence between the two phenomena under consideration is expressed through a regression line.
The regression line in a simple linear regression is expressed by the following equation:

$$y = \alpha + \beta x + \epsilon \tag{3.1}$$

where:

y is the value of the dependent variable;
α is the intercept of the regression line;
β is the slope of the regression line, so it indicates how much varies in average y to the unitary variation of x;
x is the independent variable;
ϵ is an error term.

The aim of linear regression models is to find the regression line that best fits the data. In order to do it, it is necessary to calculate the values of the regression coefficients α and β. To calculate these two coefficients the method of Ordinary Least Squares (OLS) is used: this method aims to minimize the sum of squares of the distances between the observed values and the points corresponding to them on the regression line.
Therefore it is necessary to calculate:

$$min \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3.2}$$

where: N is the number of observations taken into account; $\hat{y}$ is the fitted value of y; i indicates the observation to which the dependent variable is referred.

Given that:

$$\hat{y_i} = \hat{\alpha} + \hat{\beta} x_i \tag{3.3}$$

Then:

$$\sum_{i=1}^{N} [y_i - (\hat{\alpha} + \hat{\beta} x_i)]^2 = \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = L(\hat{\alpha}, \hat{\beta}) \tag{3.4}$$

Minimizing the function $L(\hat{\alpha}, \hat{\beta})$ means finding the values of the regression coefficients $\hat{\alpha}$ and $\hat{\beta}$ such that:

$$L(\hat{\alpha}, \hat{\beta}) \leq L(\alpha, \beta) \;\; \forall \alpha, \beta \in \mathbb{R} \tag{3.5}$$

In order to do this it is necessary to calculate the partial derivatives of the function $L(\hat{\alpha}, \hat{\beta})$ with respect to the coefficients $\hat{\alpha}$ and $\hat{\beta}$ and equal them to 0:

$$\begin{cases} \frac{\partial S}{\partial \hat{\alpha}} = -2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \\ \frac{\partial S}{\partial \hat{\beta}} = -2 \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0 \end{cases} \tag{3.6}$$

From this it is obtained:

$$\hat{\beta} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad \hat{\alpha} = \bar{y} - \beta \bar{x} \tag{3.7}$$

Once the coefficients have been calculated, it is necessary to determine whether they have a sufficient degree of significance. Taking into account the value β (the slope of the regression line) we will proceed by doing a significance test that can be of three different types.
The first possible significance test is the two-sided hypotheses test:

$$\begin{aligned} H0 &: \beta = 0 \\ H1 &: \beta \neq 0 \end{aligned} \tag{3.8}$$

The null hypothesis ($H0$) indicates that the slope parameter β of the regression line is null and therefore the independent variable has no influence on the dependent variable: in this case the regression line is a line parallel to the x-axis which is not able to explain the relationship between the two variables. The alternative hypothesis ($H1$) indicates that the slope parameter β has a non-zero value and therefore if the independent variable varies the dependent one varies too.

In addition, it is established a level of significance called α which defines the area of acceptance (and rejection) of the null hypothesis: in the case of two-sided test the area of acceptance is equal to $\frac{\alpha}{2}$ in the left end and $\frac{\alpha}{2}$ in the right end. The conventional value of alpha is 0.05 but 0.01 and 0.10 are also commonly accepted.

The next step is to calculate the test statistic as follows:

$$\text{t-test} = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})} \tag{3.9}$$

where: β^* is the value of β under the null hypothesis; $SE(\hat{\beta})$ is the standard error of $\hat{\beta}$.

Then the statistical tables of t-distribution are used to find the critical value (t-crit) related to the significance level and the degree of freedom. The value of the test statistic (t-test) is compared to the latter to see whether it is in the rejection area or not:

- If $|t - test| \leq |t - crit|$ then H0 is not rejected;
- If $|t - test| > |t - crit|$ then H0 is rejected.

The second possible significance test is the right-sided hypotheses test:

$$\begin{aligned} H0 &: \beta = 0 \\ H1 &: \beta > 0 \end{aligned} \tag{3.10}$$

In this case, the rejection area is located at the right end of the probability distribution and is equal to α.
The test statistic is compared to the critical value in the following way:

- If $t - test > t - crit$ then $H0$ is rejected;

- Otherwise, $H0$ is not rejected.

The third possible significance test is the left-sided hypotheses test

$$\begin{aligned} H0 &: \beta = 0 \\ H1 &: \beta < 0 \end{aligned} \tag{3.11}$$

In this case, the rejection area is located at the left end of the probability distribution and is equal to α and the test statistic is compared to the critical value in the following way:

- If $t - test < t - crit$ then $H0$ is rejected;

- Otherwise, $H0$ is not rejected.

Finally, the significance of the parameter β can be estimated by comparing the p-value with the value alpha of statistical significance. P-value measures the likelihood relative to the test statistic: it indicates the probability that data taken into consideration occur under the null hypothesis. The comparison between the p-value and the level of significance works in the following way:

- If $p - value < \alpha$ then $H0$ is rejected, so there is statistical significance in the linear model;

- If $p - value \geq \alpha$ then $H0$ is not rejected, so there is not statistical significance in the linear model.

At this point it is important to determine how well the regression line can explain in a significant way the relationship between the dependent variable y and the independent variable x.
For this purpose it is necessary to implement a decomposition of the Total Sum of Squares (TSS), by distinguishing the Residual Sum of Squares (RSS) and the Explained Sum of Squares (ESS).
The decomposition of the Total Sum of Squares (TSS) is shown below:

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 \tag{3.12}$$

where: $\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 = TSS$;
$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = RSS$;
$\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2 = ESS$.

From Equation 3.12 it is possible to observe that the smaller RSS, the more the regression line succeeds in explaining the analyzed data. Consequently, this result is obtained if the total variance of the dependent variable (TSS) is predominantly composed of the variance of the theoretical values (ESS).

In order to determine how well the linear model explain the relationship between the dependent variable and the independent variable, the coefficient of determination R^2 is used:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \quad where\ 0 \leq R^2 \leq 1 \tag{3.13}$$

From Equation 3.13 it is possible to conclude that if $RSS = 0$ then $R^2 = 1$ and therefore the linear model perfectly explains the relationship between the independent variable and the dependent variable. Instead if $ESS = 0$, then $R^2 = 0$ and therefore the model is not able to explain the relationship between the independent variable and the dependent variable; in this case the regression line is parallel to the x-axis, because $ESS = 0$ and therefore $\hat{y}_i = \bar{y}$.
These two represent the extreme cases, but regarding intermediate values

between 0 and 1 the linear model has a higher degree of statistical significance when R^2 is closer to 1.

3.2 Multiple Linear Regression

Multiple linear regression can be thought of as an extension of simple linear regression. In this case instead of a single independent variable, multiple independent variables are taken into account. The aim of the multiple linear regression is again to describe and evaluate a relationship between the dependent variable y and the independent variables.

The equation describing a multiple linear regression is as follows:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt} + \epsilon_t \tag{3.14}$$

where:
y is the dependent variable;
x_{kt} are the independent variables;
t indicates the observation to which each independent variable is referred;
k is the number of independent variables;
β_0 represents the intercept of the regression line;
β_k (with $k > 0$) are the parameters that quantify the effect of each independent variable on the dependent variable, considering all the other independent variables constant;
ϵ is an error term.

In multiple linear regression it is fundamental to properly choose both the number of independent variables and which of them to consider. The addition of independent variables causes a change in the linear model not only as regards the influence on the response variable, but also on all other coefficients of the regression line. With regard to the right choice of variables, those with redundancy should be avoided. Redundancy occurs when two independent variables have the same influence on the dependent variable: it is difficult to separate precisely the two influences. Redundancy leads to the problem of multicollinearity, in which two independent variables are highly correlated with each other and therefore they have the same influence on the dependent variable.

Also in multiple linear regression the coefficient of determination R^2 is used to determine the significance of the model:

$$R^2 = \frac{ESS}{TSS} \tag{3.15}$$

where:
$ESS = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$;
$TSS = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$.

In this case, however, the measure of R^2 may encounter some problems: as the number of independent variables increases, whether they are significant or not, the coefficient of determination increases in value. So with a higher number of independent variables the model seems to have a higher degree of statistical significance, but this may not indicate the truth.

To remedy to this problem it is necessary to introduce another coefficient, called *Adjusted* R^2 (R^2_{adj}):

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(N-1)}{N-k-1}\right] \tag{3.16}$$

where: N is the number of observations; k is the number of independent variables. This coefficient is able to determine the statistical significance of the model: it is not altered by the number of independent variables.

The concepts described so far regarding linear regression models are used to perform the analyses described in Chapter 5. In those analyses, in order to understand the significance of the model, we have reported the following statistical measures:

- The p-values relative to each estimated parameter, that are useful to tests the null hypothesis that the predictor has no correlation with the dependent variable (having no effect on the regression): low p-values (< 0.05) lead to reject the null hypothesis, indicating that variations in the predictor's value are related to variations in the response variable;
- The coefficient of determination R^2 and *Adjusted* R^2;
- The p-value of the F-test, that represents the probability to obtain a F-statistic value greater than the F-value of the model, under the null hypothesis that the regression is not significant [14].

Chapter 4

An Introduction to Artificial Neural Networks

Today's computers are able to perform complex calculations and solve problems that until a few decades ago seemed unattainable. Scientific research and our daily life have undergone a significant improvement thanks to the introduction of modern computers that, despite the technological evolution of recent decades, are still inspired by the architecture defined by Von Neumann [125] in the last century. However, they are not yet able to solve problems that are instead faced by man without any particular difficulty. For instance, computers are capable of performing very complex calculations quickly, but unlike the human mind, they are not capable of making assumptions or coming to conclusions about them.

Solving problems that need subjective evaluation requires human interaction with the computer, which uses a symbolic representation and processing of knowledge. This interaction and the subsequent operations of the computers are based on languages defined by precise formal rules, according to which it is possible to ensure that a sequence of symbols has a meaning; the representation of data consists of a sequence of symbols to which it is possible to give a meaning according to defined rules.

On the contrary, in intelligent life forms knowledge is distributed within the system. The brain does not use mechanisms for the representation of symbols, and operations take place on the elements that compose it, without them being in correspondence with *something* outside it.

Often it is impossible to give a computer all the information it needs to be able to deal with a complex problem, for the resolution of which previous knowledge is necessary. Starting from these limits, many researchers have felt the need to develop new computing paradigms inspired by the neurophysiological functioning of the human brain. This is how Neuro-Computation was born, that is, the discipline that, taking its inspiration from the rules of neuro-biology, allows to effectively infer the relationships that may exist between input variables and output variables. This led to the birth of neural networks which, through alternating fortunes, have been applied to a wide variety of fields, with noteworthy performance in problems such as classification, filtering, pattern association, optimization, conceptualization and prediction.

4.1 Neural Networks

Neural networks are high-level algorithms computational tools that are inspired by the functioning of the human brain; they are composed of processing elements operating in parallel that, taken individually, are able to perform simple operations: they integrate the information coming from other elements by performing an activation function and they communicate the result of this processing to other elements to which they are connected. The interaction of these simple operations via distributed elaboration leads to the execution of very complex tasks. These processing elements are inspired by biological neurons, of which, however, they represent a strong simplification.

4.1.1 The main idea

Within the human brain, it is possible to identify a large number of processing units: neurons. They are composed of three regions: the cell body (soma), the dendrites, and the axon.

The cell body contains the nucleus of the neuron and it is covered by a membrane containing channels that allow communication between the inside and outside of the soma. Dendrites represent the input channels of the neuron: they receive signals from the neurons to which they are connected. The axon instead is the output channel: its length can reach a great distance from the cell body, and it represents the pathway through which the signal emitted by the neuron propagates to other parts of the nervous system, even very remote.

Information transfer from axon to dendrite occurs in highly specialized contact zones: synapses. Each neuron may have a number of synapses ranging from a few hundred to several thousand. The information transmitted by the soma consists of an electrical signal that varies from a dozen (in a resting state) to about 500 pulses per second. The transmitted signal is higher the more the neuron is excited. This signal (action potential) starts from the soma and runs along the axon, until it reaches the presynaptic termination. Here, the action potential allows the release of a neurotransmitter that propagates through the synapse and reaches the postsynaptic dendritic zone, generating a new electrical signal that is combined with that of other dendrites and transmitted along the dendritic tree. The signals can have different effects depending on whether the synapse is excitatory or inhibitory, and their combination causes a change in the so-called “membrane potential" of the receiving soma. If this exceeds a certain bias the neuron generates a new bioelectric signal that is transmitted by the same mechanism to the neurons to which it is connected, otherwise no signal will be transmitted.

In this mechanism, a special importance is given to synapses, since the effectiveness of signal transmission, called synaptic bond strength, varies from synapse to synapse. The signal received will therefore depend on the impulses transmitted by other neurons and the strength of the synaptic bond. As early as 1949, Hebb [57] demonstrated that this force is subject to change, revealing that learning is due to synapses. However, modification of synaptic bond strength can also occur temporarily, and current research is directed toward identifying the factors by which this modification can occur. We can here summarize the main features of the brain's processing system:

- The electrical impulse transmitted by the neuron runs at a speed of 130 meters/second;
- In the brain, there is a number of neurons ranging from a few hundred billion to a few trillion;
- The density of neurons is about 80000 neurons/square millimeter and connections may be present between neurons far apart;
- The groups of neurons process information simultaneously: that is, we have parallel processing. This leads to the emergence of cognitive processes;
- Knowledge is distributed throughout the network and the evolution of the brain structure is continuous;
- The network is robust with respect to failures: the malfunction of a few neurons does not affect the overall functioning of the brain, with respect to which there is only a decrease in performance.

4.1.2 Artificial Neural Networks

Artificial Neural Networks (ANN) are inspired by the behavior of the brain, of which they represent a simplification. They are composed of elementary

units (i.e., neurons) and weighted oriented hedges to connect them (i.e., the synapses). Each neuron is associated with a numerical value that represents the value that will be transferred from the neuron and that depends on the input signals transmitted by the synapses, an activation function and an output function. Synapses are also associated with a number, which determines the magnitude of transmission.
Neurons can be classified, in relation to their function, into three categories:

- input neurons, that is, the neurons whose activations represent the input values of the network;
- output neurons, that is, the neurons whose activations represent the output of the network;
- hidden neurons, the remaining neurons, so called because they are not visible from the surrounding environment.

The behavior of a neural network is determined by:

- the activation function, which determines the output value of the neuron from the activation of the neurons connected to it;
- synapses, which determine the amount of activation of the neuron that is transferred to the neurons to which it is connected;
- network topology;
- the temporal dynamics, which determines when to update the activation values of the different neurons and the criteria for updating them (whether all neurons must be updated simultaneously or whether only some of them must be updated, and in the latter case how to choose the neurons to be updated).

Factors that characterized a network can be determined a-priori by the user or by the network itself. Generally a network is built in a hybrid way, in which some characteristics are defined by the developer while others are subjected to training. Finally, it should be noted that generally training algorithms are developed to be applied to a particular network architecture, so the choice of a particular algorithm often strongly influences the choice of the architecture and the temporal dynamics of the network.

4.1.3 The binary threshold neuron and the activation function

A neural network can be seen as a graph composed of nodes (neurons, proposed by McCulloch and Pitts [88] in 1943 and called *Processing Elements*) and connections (synapses) weighted and oriented between them. The neuron is characterized by an input operator, which determines the summation of the input values, determined by the output of the neurons connected to its input, each multiplied by the weight w of the respective synapse: the summation determines the postsynaptic potential of neuron j ($net(j)$).

$$x_1, ..., x_n \tag{4.1}$$

$$net(j) = \sum_i y_i w_{ij} \tag{4.2}$$

This value is then processed by an appropriately defined function. The first to be used was the Heaviside function, which compares the value of the summation with a *threshold* (or bias) to produce the output value: if the postsynaptic potential is less than or equal to the threshold the output of the neuron will be zero, otherwise it will be one.

$$y_j = \begin{cases} 0 & \text{if } net(j) \leq \theta \\ 1 & \text{otherwise} \end{cases} \tag{4.3}$$

In this case it is considered that the output value of the neuron can assume value 0 or 1 (step function), but it can also be established that this assumes value -1 or 1 (sign function).

$$y_j = \begin{cases} -1 & \text{if } net(j) \leq \theta \\ 1 & \text{otherwise} \end{cases} \tag{4.4}$$

This function is very simple to compute but it is not differentiable.
Other functions are used today and the most widely used of these is the logistic or sigmoid function:

$$y_j = \frac{1}{1 + A^{net(j)-\theta}} \tag{4.5}$$

the values of it belong to the interval (0,1) and it is able to reduce the possible interference of the outliers. The choice of the value A is arbitrary, but it is generally made to correspond to e ($e = 2.718283$) and this is the

value that it will assume from here on. It should be noted that the choice of A value affects the slope of the function, which converges to the step function for very large values of A and to the linear function for very small values of A.
Another widely used function is the hyperbolic tangent function, which allows the output to take values in the range (0,1).

$$y_j = tanh(net(j) - \theta) = \frac{(e^{net(j)-\theta}) - (e^{-(net(j)-\theta)})}{(e^{net(j)-\theta}) + (e^{-(net(j)-\theta)})} \tag{4.6}$$

Although there is no generalized rule for determining what function to use, it is believed that it must be continuous, differentiable, and nonlinear. This last requirement would seem to exclude the use of the linear function

$$y_j = a \times net(j) + b \tag{4.7}$$

that instead has found some applications in the output units of the network, as it avoids that the result tends towards the minimum or the maximum. Its use in hidden neurons instead is at least inappropriate, because it would determine a connection based on the type of function that is desired to avoid with the use of the neural network.
The case of the linear threshold function is different:

$$y_j = \begin{cases} 0 & \text{if } net(j) < 0 \\ 1 & \text{if } net(j) > 1 \\ net(j) & \text{otherwise} \end{cases}$$

it can assume values in the range [0,1].

A number of functions that calculate the activation of the neuron based not only on the chosen function but also on the previous activation of the neuron have been proposed. The purpose of the activation function is to bring the output back within a predetermined range, which is typically [0,1] or [-1,1], otherwise its value may assume values that are too large.

It is necessary to make a clarification about the nature of activation functions: they can be classified into deterministic functions and stochastic functions. In the former, the activation of the neuron is calculated deterministically as a function of synapse weights, input values and, possibly, the threshold and previous activation. In the latter, activation is a probabilistic function of current activation and synapse weights.

4.2 Related works

4.2.1 The birth of Artificial Neural Networks

The date of birth of the study of neural networks is conventionally traced back to 1943, when McCulloch and Pitts [88] published a paper in which they made assumptions about the method of computation used by the brain. In this study, the neuron was represented as a two-valued logic decision element with threshold activation. In addition, a first neural architecture was hypothesized, in which synapse weights were fixed. This was followed by the work of Hebb [57], who argued that the training is due to a change in the transmissive efficiency of synapses: if two neurons are active simultaneously, the efficiency of the synaptic connection between the two neurons increases.

From this explanation derive the two so-called "Hebb's rules": the presynaptic rule and the postsynaptic rule. According to them, the value of a synaptic connection is increased every time the presynaptic unit (i.e., the one that sends the signal) and the postsynaptic unit (i.e., the one that receives the signal) are active. On the other hand, the mechanism of decrease is handled differently: while in the presynaptic rule the weight of the connection is decreased if the presynaptic unit is active and the postsynaptic unit is inactive, in the postsynaptic rule this is decreased if the postsynaptic unit is active and the presynaptic unit is inactive.

4.2.2 Two layers networks and modifiable synapses

It is important to consider then the research conducted by Rosenblatt in 1962 [109], in which it was proposed a neural network with two layers (one input and one output) of threshold units (such as those of McColluch-Pitts) and with modifiable synapses in which each input unit is connected to all output units. The proposed model, called Perceptron (treated below), could be trained to classify a set of instances based on their similar characteristics. Thanks to Rosenblatt, the first theorem of convergence of the weights to values that guarantee the correct response of the network was obtained.

Similar to Perceptron was the Adaline (treated below) of Widrow and Hoff (1960) [59]. The researches of these authors were important because they offered concrete results about the practical applications and the generalization capacity of neural networks.

However, soon great limitations emerged: in 1969 Minsky and Papert [91] showed that Perceptron and Adeline were not able to distinguish a T from a C and that they were able to solve only linearly separable problems. They showed that the convergence process was too slow, that the number of logic circuits needed was sometimes too large and they expressed total distrust about further research in that field.

The prominent position of these authors in the international scientific scene froze for a certain period the studies in this field, also due to the fact that the US government decided not to renew the funding (see Floreano, Nolfi [53]). It was necessary to wait until the early 1980s to arrive at new developments on the research.

4.2.3 Nonlinearly separable problems

In 1982 Hopfield [113] proposed a completely new model of neural network based on neurons completely interconnected to each other but without self-connections, whose behavior can be understood by comparing it to a dynamical system. In this network each neuron can act as input and output. The weights are calculated deterministically and do not vary during the phase of use. When a pattern is presented at the input, the network determines a certain output that will be presented again at the input. In this way the network evolves autonomously, until it reaches a state in which the output is equal to the input. The limit of this model resides in the limited number of storable patterns (0.138 times the number of neurons). However, it has found discrete application as a self-associative memory for the recognition of corrupted configurations, the recovery of missing information and for solving optimization problems.

A few years later (Hinton and Sejnowski in 1987 [62]; Ackley, Hinton, and Sejnowski in 1988 [2]) a model was proposed to improve the Hopfield network by introducing hidden neurons and a stochastic output function. This model was called the Boltzmann Machine and is capable of solving nonlinearly separable problems. Having hidden neurons allows the network to greatly increase its storage capabilities, this network has a large flexibility, but the training procedure is very complicated and requires a very long time: for this reason it has a strongly limited use.

At the same time as Hopfield's work, Kohonen (the first publication was in 1982, followed by others in 1989 [75] [76]) proposed a completely new neural network model, which represents one of the most important European contributions to the research on neural networks, and which is the first to propose a complete system of unsupervised training (discussed below). This network consists of an input layer and an output layer. The neurons of the two layers are completely connected to each other, while the neurons of the output layer are connected in a way that they are organized on a line or a plane (usually as a matrix). In this way each output neuron is connected with a “neighborhood" of neurons and learning follows a logic of competitive type, where the weights afferent to the neurons of the neighborhood of the winning output unit undergo a modification of the weights in relation to a "Mexican hat" function. This network, thanks to its peculiarities, could be defined biologically plausible and suitable for application in many areas such as motor control and speech recognition.

4.2.4 Feed-forward networks

The work of Rumelhart, Hinton and Williams in 1986 [110] was essential for the development and application of neural networks. It proposed a learning algorithm for multi-layer feed-forward networks that overcomes the limitations highlighted by Minsky and Papert. This was the Back-Propagation algorithm (treated below), which would soon establish itself as the most widely used learning algorithm for neural networks. This algorithm systematically modifies the weights of the synapses, making the response of the network get closer and closer to the desired one provided by the user. Thanks to this work, the scientific community's attentions turned with renewed vigor towards neural networks, as it provided an easy to implement method capable of providing good performance.

Finally, it is worth mentioning the work of Jordan (1986) [67] and Elman (1990) [49] who proposed variants of a normal Feed-Forward network, adding connections from an upper to a lower layer and self-connections on one or more nodes: recurrent topology networks.

In recent years, research has led to the development of new learning algorithms, the most interesting of which are those that are based on genetic algorithms, and new analysis tools to better understand the behavior of various neural network models. Many researchers today make use of neural networks in their respective research fields, and there are several applications available on the market that are based on them. This last fact is also due to the current availability of low cost computing resources, a phenomenon that for several years has led to the use of software simulators of neural networks, diverting interest from their physical implementation.

4.3 Main characteristics of neural networks

Neurons of the type seen above (*Processing Elements*) can be organized into different architectures to form a neural network.

A first architecture proposes neurons that are fully connected to each other, resulting in a fully connected structure (e.g. the Boltzmann machine seen above).

Another architecture proposes neurons grouped in different layers, conceived as a disjointed and ordered subset, according to their function: there is therefore an input layer, one or more hidden layers and an output layer. To each layer neurons of adjacent layers and, eventually, of the same layer are connected.

The most proposed and examined model involves each neuron being connected to all neurons in adjacent layers and no connections between neurons in the same layer. Neurons in the input layer have no input connections and their activation consists of the pattern corresponding to a problem input.

A function transfers the activation value without performing calculations to the neurons of the hidden layer, which calculate their activation and transfer it either to the neurons of another hidden layer or to the output neurons. The activation of these represents the output of the network. The information flow of the network is unidirectional: the neurons receive input only from the previous state and transmit it only to the next one. For this characteristic this architecture is called feed-forward.

There are also architectures with partial connections, in which each neuron is connected only with some other neurons of the adjacent layers.

Furthermore, the recurrent networks represent a variant of the layered structure just described. These are feed-forward networks to which extra neurons are added. These neurons can be *state neurons* or *context neurons*; the first ones indicate temporal memory: two equal inputs can correspond to two different outputs depending on the temporal sequence. The second ones can be equated to input neurons, but instead of receiving signals from outside, they receive them from the network itself.

Discussing about neural networks, the learning process has to be mentioned: it is the ability of the network to modify its behavior in order to obtain the right output given some input. Usually, this process occurs mainly through the modification of synaptic connections (weights), and this led to the development of training algorithms that became common and led to an increase in the spread of neural networks (e.g. Back-Propagation [110]).

There are three main types of learning: *supervised learning, unsupervised learning* and *reinforcement learning.*

In the supervised learning [25] it is supplied to the network a training set, formed by a set of inputs to which correspond certain outputs. The network, through their analysis, learns the possible existing relationships between input and output; in this way the network learns to generalize the same relationships and then it is able to calculate new ones to be applied to unknown inputs. While the network elaborates the outputs, the weights of the synapses are modified in order to obtain correct answers: those that determine the correct outputs are increased and those that generate invalid values are decreased. There is a supervisor able to determine if the output is right or wrong and the network is able to determine the error, given by the difference between a desired value and an actual value. In this process the experience of the user is crucial, as he has to provide a training set of dimensions and input-output relations suitable for the network.

In unsupervised learning [13], on the other hand, only a set of inputs is provided to the neural network. Through the analysis of it, the network creates representative clusters to be able to categorize them. Again, in order to obtain the correct outputs, the weights of the synapses are modified, but now it is the nodes of the network themselves that modify them.

Finally, in reinforcement learning [116], no sets of inputs or outputs are provided to the network, nor is the goal to obtain correct outputs through various changes in synaptic weights. The neural networks, in fact, learn exclusively from the interaction with the environment: they go towards the desired result through incentives (positive actions) and disincentives (negative actions). The reinforcement is represented by the action that allows to approach the result.

4.3.1 The Perceptron

The Perceptron is a model of a neuron proposed by Resenblatt in 1962 [109], characterized by binary input, binary output, and a Heaviside activation function. This neuron is the simplest example of neural network and has found immediate application in the recognition of forms offering good results. The main difference with the neuron of McCulloch and Pitts (discussed above) is that the weights are modifiable depending on the input-output association that is desired from the network.

The limitation of the Perceptron, however, is that it cannot solve problems characterized by non-linearly separable inputs: a problem is said to be linearly separable if, arranged the inputs in a space, it is possible to determine a hyperplane that clearly divides the class of inputs leading to a positive solution from those leading to a negative solution.

It is useful to describe the functioning by thinking of a two-input classification problem x_1 and x_2. The Perceptron receives the input pattern and weights its components through connection weights to determine whether or not the pattern belongs to a particular class. Given the connections w_1 and w_2, it is possible to separate the reference plane with the line

$$w_1x_1 + w_2x_2 = \theta \tag{4.8}$$

which separates the plane into two half-planes for which we have:

$$w_1x_1 + w_2x_2 > \theta \tag{4.9}$$

and therefore $y = 1$ in the half-plane to the right of the line;

$$w_1x_1 + w_2x_2 \leq \theta \tag{4.10}$$

and thus $y = 0$ in the half-plane to the left of the line.

It is useful to introduce a dummy neuron with activation equal to -1 and synapses equal to the threshold value into this scheme. The training of the Perceptron is done in a supervised way. If the desired output and the actual output do not coincide, the synapses will be modified according to the rule[6].

$$w_i(t+1) = w_i(t) + \eta\delta x_i \tag{4.11}$$

where:

$w_i(t+1)$ is the synapse weight i after the modification;
$w_i(t)$ is the synapse weight i before the modification;
η is the learning rate;
x_i is the transmitted value by synapse i;
δ is the difference $(d - y)$ between desired output d and actual output y.

The meaning of this formula can be easily understood: if the neuron receives as input a pattern that does not belong to the class object of the application, it can return an output value of 0, and in this case the output is correct, or 1. In this second case it is verified that the value of the summation $w_1x_1 + w_2x_2$ is too large, so it must be decreased by modifying the value of the synapses, which will have a new value lower than the previous one.

On the contrary, if the neuron receives an input pattern belonging to the class it can return in output the value 1, and in this case the output is correct, or 0. In this second case it is verified that the value of the summation $w_1x_1 + w_2x_2$ is too small, so it must be increased by modifying the value of synapses, which will result to have a new value higher than the previous one.

According to what has just been said, the modification of synapses occurs only in case of misclassification of the pattern and does not take place

[6]This rule is called Delta-rule.

if the output of the presynaptic neuron is null. In this example the effect of learning is to change the inclination of the line that divides the plane in two until the separation is achieved.

The parameter η determines the speed of learning of the network: high values lead to large changes in the synapses at each step with possible learning instability, low values lead to small changes.
The learning algorithm can be schematized as follows:

- a set of (X, D) pairs of available examples is determined;
- one initializes the weights w with random values;
- a pair (x, d) is presented;
- the response y of the network is calculated;
- if the effective result y and that wished D do not coincide the synapses are modified in base to the Delta rule, otherwise they remain unchanged;
- one presents a new couple and one proceeds as above, until the exhaustion of the available examples.

The Delta rule was later generalized by Widrow and Hoff [59] in 1960. They applied it to bi-layer networks composed of ADELINE *(Adaptive Linear Neuron, Adaptive Linear Elements)* with multiple output units, able to discriminate different linearly separable classes: each output unit corresponds to a class and when a pattern belonging to a class k is present, the network determines the activation of the node o_k corresponding to that class.

These two authors introduced the concept of error, by assessing that the weights variation is proportional to the error gradient. The absolute quadratic error relative to an input pattern is the summation of the squares of the differences between actual output and desired one, relative to each output node of the network.

$$E = \frac{1}{2}\sum_{o=1}^{m}(d_o - y_o)^2 \tag{4.12}$$

This can be minimized by modifying the connection weights: if the error increases when the weights w increase, then the derivate of E respect to w has positive sign; so the weights have to be reduced. Otherwise, if the error decreases when the weights w increase, then the derivate of E respect to w has negative sign; so the weights have to be increased. In detail, the

Widrow-Hoff rule explains that the weight variation must equal the inverse of the derivate of the error respect to the weight and multiplied by the learning rate:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \tag{4.13}$$

It is possible to apply the rules of formal derivation as follows:

$$\frac{\partial E}{\partial w_{ij}} = (y_j - d_j)\frac{\partial y_j}{\partial w_{ij}} = (y_j - d_j)\frac{\partial y_j}{\partial net_j}\frac{\partial net_j}{\partial w_{ij}} \tag{4.14}$$

Since

$$\frac{\partial y_j}{\partial net_j} = f'(net_j) \quad \text{and} \quad \frac{\partial net_j}{\partial w_{ij}} \tag{4.15}$$

it will be obtained

$$\Delta w_{ij} = -\eta(y_j - d_j)f'(net_j)x_i \tag{4.16}$$

and since

$$\delta_j = (y_j - d_j)f'(net_j) \tag{4.17}$$

the final formula will be

$$\Delta w_{ij} = -\eta\delta_j x_i \tag{4.18}$$

This mechanism could be interpreted by saying that the error is a quadratic function of the network weights, and the learning algorithm goes down along the line with the maximum slope of the function, starting from the generic point determined by the initial weights randomly selected. The learning rate represents the pace length of the decline.
Starting from this rule, Rumelhart and others developed the Back-Propagation algorithm.

4.3.2 Learning algorithms

As it has been already treated, Perceptron is able to provide good performance but it is not able to deal with non-linearly separable problems. To explain this type of problem, the example of XOR function (exclusive or) can be used: it is not possible to draw a line on the plane in order to divide the solution space in two different categories.

Non-linearly separable problems, such as the one described above, could be solved by applying a multilayer feed-forward architecture which has one or more hidden layers. This kind of network has been initially called MLP (*Multi-Layer-Perceptron*: this name is not suitable nowadays). It resulted limited firstly because Perceptrons are able only to solve linearly separable problems; then, because there were not learning algorithms which could be eligible: the learning algorithm Widrow-Hoff is impossible to apply due to the introduction of hidden layers. The algorithm is based on the modification of the weights depending on the difference between the actual and desired output per unit and since it is not possible to determine the desired output of the intermediate units, then it is impossible to apply the algorithm to this type of network.

In 1986 Rumelhart, Hinton and Williams [110] proposed the Back-Propagation algorithm, which is based on the back-propagation of the error from the output units to the input ones. This algorithm can be divided into two phases: in the first one, the *forward phase*, the activation of the input units is propagated forward through the activation functions. In the second one, the *backward phase*, the weights of the connections are modified through the technique of the decline of the gradient; with this technique the error of the output units is propagated backward until the input units.

The error of the hidden units is not calculated in the usual way as for the output units, but by multiplying the error of the output units, weighted by the relative weight of the connection, by the first derivative of the output function of each hidden unit: this is the major difference in the algorithm.

$$\delta_j = \begin{cases} f'(net(j))(y_j - d_j) & \text{if j is an output unit} \\ f'(net(j)) \sum_k \delta_k w_{jk} & \text{if j is an hidden unit} \end{cases} \tag{4.19}$$

$$\Delta w_{ij} = -\eta \delta_j x_i \tag{4.20}$$

This way the weights are modified basing on their influence on the error of that example.
It is known that

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net(j)} \frac{\partial net(j)}{\partial w_{ij}} \tag{4.21}$$

$$\frac{\partial y_j}{\partial net(j)} = f'(net(j)) \tag{4.22}$$

$$\frac{\partial net(j)}{\partial w_{ij}} = x_i \tag{4.23}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} f'(net(j))x_i \tag{4.24}$$

To obtain $\frac{\partial E}{\partial w_{ij}}$ it is needed $\frac{\partial E}{\partial y_j}$. So it is needed to express the intermediate nodes errors in terms of output units errors, which are the only ones known.

$$\frac{\partial E}{\partial w_{ij}} = f'(net_j)x_i \sum_k (\delta_k w_{jk}) \tag{4.25}$$

so

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = -\eta \delta_j x_i \tag{4.26}$$

This algorithm can be applied to networks with any number of layers, although it has been shown that networks with more than four layers do not have higher computational power. Particular attention should be given to the selection of the activation function: the sigmoid one is the most used, but there are no codified indications in this regard. There is only one rule: do not apply linear activation functions on all neurons; this would result in a network with any number of layers functioning as a two-layer network, since the combination of linear functions is itself a linear function.

In a multilayer network each hidden unit identifies a hyperplane which separates the pattern space into two distinct classes. Thanks to the combination of the hyperplanes it is possible to perform quite difficult classification tasks and obtain good results. Each hidden neuron, in fact, can be trained to activate in case of a certain feature in the input pattern: this is useful to model neurophysiological features and cognitive processes.

The term η has a particular importance: it represents the learning rate that determines the learning speed of the network, as already seen in the previous algorithms. If the learning rate is small, the training time may be too long; on the contrary, if the learning rate is very large, it is likely that oscillatory behaviors occur in the network around the minimum point of the error function. Therefore it is useful to find a compromise in the following way: to use a great rate of learning at the beginning of the process, in order to accelerate the convergence, and then to reduce it gradually in order to avoid oscillations at the end of the process.

The last formula can also be improved by adding another term, called *momentum*, which represents a proportion of the last change made to the weight:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} + \beta w_{ij} \tag{4.27}$$

This variant was already proposed by Rumelhart in the article in which the Back-Propagation [110] was proposed for the first time. The global error can be thought as a function of the synaptic weights, characterized by a very irregular trend due to the non-linearity of the activation of the output units. The Back-Propagation algorithm performs a search for the minimum in this surface, decreasing at each iteration the value of the error. The addition of the *momentum* term, therefore, ensures that the error surface is crossed quickly with few steps in the presence of plateaux, while the step size decreases as the surface becomes irregular. In this way, even with a high learning rate there is no risk of large oscillations because the network recovers a good portion of the last weight reached.

In the process of supervised learning with the algorithm of Back-Propagation, a couple of samples is furnished to the net at each iteration:

$$X = (x_1 ... x_n) \quad D = (d_1 ... d_n) \tag{4.28}$$

They represent the input and the response that is desired to be produced by the network for that particular input. When an input is provided to the network, the postsynaptic potential of the input neurons will be propagated, through the activation and transfer functions, to the output neurons, which will determine an output value y. According to this, the synapses will be modified to minimize the difference between the obtained outputs (y) and the desired ones provided by the user (d); to obtain this result, a measure such as the summation of squared errors has to be minimized.

Learning occurs on a set of samples (X, D) called the *training set*. The presentation of all the elements belonging to the set is called epoch: usually more than one epoch is necessary for the training to be concluded. Once the network is trained, the weights become unmodifiable and the use of the network on a previously unseen set of data, the *generalization set*, can begin.

One potential problem is the lack of generalization capability of the network. In this case, by minimizing the error, the network is able to provide the desired output for the training set data; however, it then commits significant errors with the unknown data. This is the phenomenon of *over-fitting* or *over-training*. One way to solve this problem is to create another set of samples (X, D) called *validation set*, used to evaluate the performance of the network on data not used for learning. In practical terms, the data belonging to the training set are used to modify the synaptic weights, and every n iterations (with n defined by the user) an element of the validation set is presented to the network only to evaluate the deviation of the output from the desired output, that is, without modifications of the weights.

Usually a set of samples (X, D), which is partitioned into training set and validation set is made available. There are no rules on how to perform this partitioning and the relationship between the two sets, but generally it is done in such a way that the number of elements present in the validation set is about one third of those present in the training set. The minimization of the error must take place on the validation set.

When the error on the validation set has an absolute minimum, then it starts to rise again, while the error on the training set continues to decrease, learning should be stopped: it represents the point at which the generalization capacity of the network is maximum. This seems to be true only theoretically, as the choice of validation set is arbitrary, and different validation sets have different error functions.

The minimum point, however, cannot be determined a priori, and this can prove to be a problem since convergence is not assured in a given time frame. In addition, the error function can be characterized by plateaux, which makes it difficult to determine if the minimum has been reached.

In order to face these problems, stop-learning criteria are used: the end of the learning phase is established on the basis of the occurrence of a condition, such as the achievement of a threshold or of a number of iterations (or epochs) without any improvement of the error. These rules indicate the priority that is given to reaching a certain level of generalizability of the network with respect to reaching the minimum of the error function.

Furthermore, if patterns are always presented in the same temporal order, there is a risk that the network will identify relationships between them only from a presentation order perspective, and this would lead to oscillatory behavior. To overcome the over-fitting, it is possible to use simple techniques like the *shuffling*.

The Back-Propagation algorithm can be divided into *online* and *offline* modes: the difference between these two lies in the way the weights are updated. In the former, a pair of values (x, d) is presented as input, then the output of the network is computed and compared with the desired one; the synapses are modified by the network according to an error metric. Subsequently, the same procedure is followed with other input patterns. In the second one, the modification of the weights is based on the weight variations calculated for each pair belonging to a certain epoch; the modification is applied only when the epoch is concluded. This can be useful when one has parallel implementations with high communication costs.

Between these two modes it is possible to identify an intermediate one: the *chunkwise Back-Propagation*. Before changing the weights, the number of patterns (*chunks*) to be presented to the network is defined. This can prove to be a useful intermediate way if it is believed that the offline mode has too

long times of convergence and the online one has excessive oscillations.
It has been shown that the online version converges with fewer iterations, but the offline version seems to have a higher probability of reaching the optimal configuration: in the latter there are no continuous sign changes in the update of the weights, so the relevance of outliers is reduced. This advantage of the offline version (also known as the *Back-Propagation store*) does not manifest itself when the online version with momentum is used: the difference reduces as the momentum increases.

Rumelhart (cited in Hanson and Pratt [56]) observed that given a set of data, the simplest and most robust network is the one that, on average, leads to the best generalization over the population from which the data were extracted. From this statement, systems have been derived to improve the performance of the Back-Propagation algorithm.

The first of these is the *reduction*: the synapses that present, in absolute value, a value inferior to a threshold, are forced periodically to assume a null value. Usually this threshold is defined as a proportion of the smallest weight within the network, but it can also be a proportion of the largest weight or of the average of the weights. The weights with null value are eliminated from the net.

With the *decay* technique, instead , a term, which represents a measure of the complexity of the network, is added to the error:

$$E = \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{r}(y_k - d_k)^2 + \lambda \sum_{ij} \frac{w_{ij}^2}{1 + w_{ij}^2} \tag{4.29}$$

Each connection will be associated with a cost as the goal is to minimize complexity. This results in a modification of the rule of learning:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} + \beta \Delta w_{ij} - \lambda \sum_{ij} \frac{2w_{ij}}{(1 + w_{ij}^2)^2} \tag{4.30}$$

it turns out to be very sensitive to the value chosen for λ; a small value has no effect, while a too large one brings all the weights to assume null value. Among the solutions there is the modification of lambda as the learning proceeds: some rules have been proposed in [37]. Another option is to reduce the weight by a proportion of the old value, as proposed by Werbos [129]. The weights are taken to assume null value if not reinforced by Back-Propagation.

$$\Delta w_{ij}(t+1) = \eta \delta_j x_i - \alpha w_{ij}(t) \tag{4.31}$$

It is possible to say that the Back-Propagation algorithm has some limitations. First of all, there is no theorem that guarantees convergence, although the error decreases with the continuation of iterations: therefore the achievement of the optimal configuration of the weights cannot be guaranteed.

The error function has a very irregular trend, so it is possible that the algorithm goes towards a local minimum and remains stuck there. One of the solutions to this problem is to periodically add a random noise to the weights in order to make sure that a little instability in the network is manifested: it is however to be treated with caution in order to avoid misleading and totally unstable behavior.

Another problem is that initializing all the weights at the same value, the changes made will be the same for all synapses that lead to the same output branch, so conventionally a random initialization of the weights between a minimum and a maximum limit defined by the user is used.

Furthermore, like the Widow-Hoff algorithm, weight updating does not occur if the output of the presynaptic unit is null.

Finally, mention must be made of the biological implausibility caused by the use of the error function and the presence of connections that behave symmetrically in that they transmit two streams of information: activation in the first phase and error in the second. Fortunately, biological implausibility has not compromised research in this direction.

4.4 Advantages and disadvantages of Neural Networks

It is possible to say that the greatest advantage of neural networks is their ability to generalize, that is, the ability to operate on data never previously analyzed, provided they have the same characteristics as those used in the training phase. This is particularly useful in prediction problems, when dealing with incomplete data.

Moreover, it is the network itself that determines the eventual relationships between input and output, without the need to establish particular a priori hypotheses in this regard: they learn from the experience and the processing of the data presented to them. The networks are able, in addition, to capture the non-linear relationships among a large number of variables.

All of these characteristics has proven to be particularly useful with regard to various economic and financial applications.

It has been particularly appreciated the ability of the networks to operate as associative memories: the network is able to go back to a pattern even when it is presented with a partial or corrupted version of it; this can be useful in speech or image recognition. A neural network is also able to deal with non-numerical data, such as geographical position.

Another advantage lies in the fact that the failure of some component of the network does not affect its overall behavior. In addition, the distributed and parallel processing typical of neural networks, and the consequent parallelism implicit in the learning algorithms, mean that they can be implemented on parallel machines, with a significant improvement in performance in terms of learning time.

A disadvantage of networks is that they are not able to explain the results achieved during training and the relationships that generate phenomena: this comes from the sub-symbolic representation of knowledge in connectionist systems.

Finally, a theory on neural networks is missing, and this has led researchers to have to develop them through try-and-error techniques. A particular problem is given by the choice of parameters, for which there are no precise rules, but that is crucial in relation to the problem of over-fitting.

Chapter 5

Our Linear Regression

In this chapter we are defining linear regression models to predict the articulation of innovation over the five sets of data defined in Section 2 (i.e., Eclipse, NASDAQ, FTSE, CAC, DAX). We are defining three different models:

- the model in which the predictor set is composed of the absolute number of male and female in the board of directors of the business taken into account: this model only considers gender aspects, and will be referred to as *GE*;
- the model in which the predictor set is composed of all co-creation components (C1-C24): this model will be referred to as *CO*;
- the model in which the predictor set is composed of all co-creation components (C1-C24) plus the absolute number of male and female in the board of directors of the business taken into account: this model considers gender and co-creation aspects, and will be referred to as *CO-GE*.

All models have been implemented in Python, which is a high-level programming language created by the Dutch programmer Guido van Rossum in the early 1990s [122]. Thanks to its dynamism, simplicity and flexibility it is considered one of the most important and widespread programming languages in technical and scientific fields. This language is mostly used for purposes such as web development and GUI programming[7], and it is useful in various fields and industries (i.e., Software Engineering, Mathematics, Data Analytics, Science, Accounting and Network Engineering). Python has become popular all over the world because of some advantages it offers. The first of

[7]Graphical User Interface is a type of user interface in which the interaction between humans and machines occurs visually through the use of graphical representations.

these consists in the fact that it is completely free, so some might imagine that it is not a suitable tool for performing specific and detailed work and that it is not often updated. The second advantage is due to the support given by a very active community, it is always kept up-to-date and it contains many libraries. There exist several web platforms devoted to the Phython community, such as *Full Stack Python* (https://www.fullstackpython.com/).

Python is pseudocompiled: the interpreter's task is to analyse the source code and execute it. For this reason Python is a portable language: it can be used on different platforms, such as Windows, MacOS, Linux, Android and iOS, provided the Python interpreter is installed.

Finally, Python is characterized by a simple and intuitive syntax, which makes it easy to learn and develop: it has different types of IDEs[8] for data science, which allow a better and easier process management of data analysis and machine learning. Among them it is worthwhile to mention *Spyder*, *Pycharm*, *Thonny*, *Atom*, *Jupyter Notebook*, *IDLE*, just to mention the most used ones. For example, using *Jupyter Notebook* enables the user to code directly in a web browser, in which both the source code and the output are displayed 'in-line'. An outline of the code used in Python to carry out the linear regression can be found in Listing 5.1.

Listing 5.1: Linear regression models Python's code

```
import pandas as pd
import numpy as np
from pandas import ExcelWriter
from pandas import ExcelFile

# Read File

df_hot_gen = pd.read_excel(r'C:\Users\saraghilardi\OneDrive\Thesis\Eclipse.xlsx')
df_hot_gen

# Variables used in the experimental phase

var_B = df_hot_gen ['customer+OR+user-dialog+OR+dialogue+OR+conversation+OR+feedback+OR+
call+OR+interact+OR+information-exchange+OR+information-sharing+OR+information-access+OR+engage']
var_c = df_hot_gen['customer+OR+user+OR+forum+OR+connect+OR+network+OR+networking']
var_d = df_hot_gen['lease+OR+rent+OR+license+OR+self-serve+OR+self-service']
var_e = df_hot_gen ['customer+OR+user-cooperate+OR+cooperation+OR+collaboration+OR+partnership']
var_f = df_hot_gen ['customer+OR+user-suggest+OR+suggestion+OR+input+OR+request+OR+demand']
var_g = df_hot_gen ['internal-expertise+OR+resource']
var_h = df_hot_gen ['customer+OR+user-risk-manage+OR+management+OR+control+OR+assess+OR+
reduce+OR+reduction+OR+potential+OR+exposure']
var_i = df_hot_gen ['customer+OR+user-IP+OR+intellectual-property']
var_j = df_hot_gen ['customer+OR+user-learn+OR+learning']
var_k = df_hot_gen ['product+OR+process+OR+service-evolution+OR+evolve']
var_l = df_hot_gen ['customer+OR+user-experience']
var_m = df_hot_gen ['customer+OR+user-test+OR+trial+OR+beta']
var_n = df_hot_gen ['integrated-online-services']
var_o = df_hot_gen ['simulation+OR+simulate+OR+model+OR+modelling+OR+virtual-world
```

[8]IDE stands for Integrated Development Enviroment. It is a coding tool that allows the user to use programming language codes in an easier way.

```
+OR+reference-design+OR+reference-flow+OR+demo-application+OR+toolkit+OR+
tutorial+OR+sdk+OR+software-development-kit']
var_p = df_hot_gen ['product+OR+process-modularity+OR+modular+OR+module']
var_q = df_hot_gen ['customer+OR+user-produce+OR+assemble+OR+manufacture']
var_r = df_hot_gen ['customer+OR+user-options+OR+choice+OR+choose']
var_s = df_hot_gen ['design+OR+process-flexibility+OR+flexible+OR+adaptable']
var_t = df_hot_gen ['customer-partnerships+OR+interaction+OR+relationship+OR+participate+OR+
participation+OR+activity+OR+action']
var_u = df_hot_gen ['cost-reduce+OR+reduction+OR+saving']
var_v = df_hot_gen ['customer+OR+user-survey+OR+review+OR+voting+OR+vote+OR+rate+OR+rating']
var_w = df_hot_gen ['trust+OR+honesty+OR+integrity+OR+transparency']
var_x = df_hot_gen ['customer+OR+user-disclose+OR+inform+OR+disseminate+OR+reveal']
var_y = df_hot_gen ['customer+OR+user-dashboard+OR+statistics']
var_z = df_hot_gen['new+AND+(product+OR+service+OR+process+OR+application+OR+
solution+OR+feature+OR+release+OR+version+OR+launch+OR+introduction+OR+
introduce+OR+new-product+OR+new-service+OR+new-process+OR+new-solution
+OR+product-launch)']
var_aq = df_hot_gen['Number␣of␣board␣male']
var_ar = df_hot_gen['Number␣of␣board␣female']

Res_hot_gen = pd.DataFrame({'var_B': var_B,'var_c': var_c, 'var_d': var_d,'var_e': var_e,
'var_f':var_f ,'var_g':var_g , 'var_h':var_h , 'var_i' : var_i ,'var_j':var_j ,
'var_k':var_k ,'var_l':var_l,'var_m':var_m,'var_n':var_n ,'var_o':var_o,'var_p':var_p,
'var_q':var_q,'var_r':var_r,'var_s':var_s,'var_t':var_t,'var_u':var_u,'var_v':var_v,
'var_w':var_w,'var_x':var_x,'var_y':var_y,'var_z':var_z ,'var_aq':var_aq,'var_ar':var_ar})
Res_hot_gen.replace([np.inf, -np.inf], np.nan, inplace=True)
Res_hot_gen= Res_hot_gen.replace([np.inf, -np.inf], 0).fillna(0)
Res_hot_gen = Res_hot_gen.astype(float)
Res_hot_gen

# Linear Regression

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn import linear_model
from sklearn.linear_model import LinearRegression
lr = LinearRegression()

# Drop the innovation metrics from the predictor set

X = Res_hot_gen.drop(['var_z'],axis=1).values
Y = var_z.values

lr.fit(X , Y)

import statsmodels.api as sm
Y = var_z.values
X2 = sm.add_constant(X)
est = sm.OLS(Y, X2)
est2 = est.fit()
print(est2.summary())
Y_pred = lr.predict(X)
```

5.1 GE model

First of all we want to introduce a model that identifies a linear relationship between the articulation of innovation and the gender components only:

$$I = \beta_0 + \beta_1 \cdot G_1 + \beta_2 \cdot G_2 + \epsilon \tag{5.1}$$

where:
I stands for the articulation of innovation;
β_0 represents the intercept of the regression line;
β_i ($i \in [1\ ,\ 2]$) is the coefficient related to the gender components ($i = 1$ refers to male; $i = 2$ refers to female);
G_1 represents the number of male on the board;

Table 5.1: Estimates of the linear model GE: estimated parameters, p-values and significance statistics of the linear model.

Estimates	Eclipse	NASDAQ	FTSE	DAX	CAC
Parameters:					
Intercept	4.8454	13.4775	-14.3853	16.9444	9.1727
G1	0.4762	-0.7576	0.7702	-0.0549	1.2245
G2	3.4124	-0.3973	4.7917	-1.8921	-1.3769
$p-values$:					
Intercept	0.352	0.063	0.121	0.156	0.630
G1	0.628	0.388	0.445	0.981	0.565
G2	0.154	0.778	**0.001**	0.700	0.472
Statistics:					
R^2	0.015	0.010	0.115	0.080	0.020
p-value of F-test	0.114	0.611	**0.004**	0.903	0.685

G_2 represents the number of female on the board;
ϵ represents an error term.

It is necessary to recall that the measures related to innovation and value co-creation have been calculated through the keywords research application described in Chapter 2. With regard to the gender, the absolute number of men ($G1$) and the absolute number of women ($G2$) in the board of directors of the businesses was either found online or asked to the business itself.

Table 5.1 shows the estimate of the parameters for each of the two gender components, for the five sets of data taken into account. In addition, in order to understand the significance of the model, we have reported the following statistical measures:

- The p-values relative to each estimated parameter, that are useful to tests the null hypothesis that the predictor has no correlation with the dependent variable (having no effect on the regression): low p-values (< 0.05) lead to reject the null hypothesis, indicating that variations in the predictor's value are related to variations in the response variable;
- The coefficient of determination R^2;
- The p-value of the F-test, that represents the probability to obtain a F-statistic value greater than the F-value of the model, under the null hypothesis that the regression is not significant [14].

For the sake of readability, p-values smaller than 0.05 are highlighted in bold.

By looking at this summary, we can see that G1 is never significant, while G2 is significant only on one (out of five) sets of data. The R^2 values are always small, and the significance of the regression is low, as assessed by the p-value of the F-test, which is always greater than 0.05 (except for the FTSE set of data). We can easily conclude that this model is not significant to explain the articulation of innovation.

5.2 CO model

As second model, we want to introduce the model in which the predictor set is composed of all co-creation components (C1-C24, as described in Chapter 2), that can be expressed as follows:

$$I = \beta_0 + \sum_{i=1}^{k} \beta_i \cdot C_i + \epsilon \tag{5.2}$$

where:
I stands for articulation of innovation;
β_0 represents the intercept of the regression line;
β_i are the coefficient related to co-creation;
C_i are the value co-creation components ($i \in [1 \ldots 24]$);
$k = 24$ is the number of co-creation variables;
ϵ represents an error term.

Table 5.2 shows the coefficients for each of the 24 value co-creation variables, along with the other measures to assess the significance of parameters and the goodness of the regression already introduced in Section 5.1. We have also reported the $Adjusted\, R^2$, that represents a modified version of R^2 that takes into account the number of predictors in the model, and can be used to compare the explanatory power of regression models that contain different numbers of predictors.

In a way, an increase of the cardinality of the predictor to a model leads the (simple) R^2 to increase, hence a model with a bigger number of predictors may show better performances (fit) due to the fact that it has more terms. The $Adjusted\, R^2$ instead penalises the introduction of useless predictors, and decreases when a predictor improves the model by less than expected by chance: its values is always equal or smaller than the simple R^2. Its formula can be expressed as follows:

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(N-1)}{N-k-1}\right] \tag{5.3}$$

where: N is the number of observation in the sample; k is the number of independent variables.

In what follows we will comment the results of the linear regression analysis by referring to the values reported in Table 5.2. The values will be described by considering one dataset at a time.

5.2.1 Eclipse

First of all from Table 5.2 we can observe the p-values related to each independent variable (C1-C24). Taking into consideration a 95% confidence interval, a variable is significant if its p-value is less than 0.05. From Table 5.2 we observe that this condition is met by the following variables: C8, C9, C14, C15, C17, C19.

Subsequently, observing the main statistics of the linear regression we can see that R^2 and $Adjusted\, R^2$ assume low values. This shows that in this dataset the co-creation variables are not able to explain the trend of the innovation variable. On the other hand, the p-value of F-test is very small and this means that there is a high statistical significance in the CO model regarding Eclipse dataset.

5.2.2 NASDAQ

As for the NASDAQ dataset, in Table 5.2 it is possible to notice that only two p-values are less than 0.05. Therefore, only two co-creation variables are statistically significant and these are: C15 and C18.

Regarding the statistics of linear regression for this dataset the values of R^2 and $Adjusted\, R^2$ are quite high. This means that the co-creation variables are able to explain the innovation variable. Also the p-value of F-test has a small value which confirms the high statistical significance of the linear model.

5.2.3 FTSE

Regarding the FTSE dataset there are some co-creation variables whose p-values are less than 0.05 and therefore are statistically significant. In partic-

Table 5.2: Estimates of the linear model CO: estimated parameters, p-values and significance statistics of the linear model.

Estimates	Eclipse	NASDAQ	FTSE	DAX	CAC
		Linear model			
Parameters:					
Intercept	-4.1437	-0.4921	-2.2563	-1.6653	-6.1916
C1	-0.5627	0.1205	-0.3374	1.2576	1.0674
C2	1.1695	0.3374	3.1604	2.8070	17.6961
C3	-1.9233	0.2098	1.3076	2.0582	9.6768
C4	-20.5127	-1.3261	0.9638	2.0582	9.6768
C5	-3.1248	1.4207	2.9320	4.0372	1.8828
C6	-3.1248	1.4207	2.9320	0.5474	-7.3992
C7	4.2589	0.1579	-0.5384	3.4252	0.9397
C8	-11.3412	-3.6831	-1.6878	-9.2725	-99.9797
C9	2.8478	1.0239	0.3229	4.1849	-16.0429
C10	-1.1368	0.0912	-2.1477	-0.5241	-4.9764
C11	-1.3556	0.0866	-0.5033	-5.1339	-5.7084
C12	1.0277	-0.6601	-13.1893	45.0063	21.0217
C13	3.2693	0.0900	1.9997	-2.3560	1.0159
C14	-1.2931	-0.1682	3.1707	0.0162	1.5252
C15	4.3344	3.0123	4.1965	-1.1954	-1.4963
C16	6.2534	1.5962	-1.4511	2.4952	8.6916
C17	-4.9335	-0.8650	-1.5943	-1.1408	1.6712
C18	0.5957	-3.8688	-7.6064	12.2532	-2.0042
C19	-6.3125	0.1737	1.9969	-10.7446	0.1459
C20	1.7789	2.1388	-0.9569	-23.5393	22.5001
C21	0.5433	1.0453	-0.0310	0.1609	-5.6271
C22	0.4984	1.1002	1.7455	-7.6088	-9.1443
C23	-15.0079	1.9771	20.6706	-59.8209	61.2465
C24	40.0454	0.5316	-6.9515	10.0600	-5.9675
$p-values$:					
Intercept	0.370	0.692	0.319	0.756	0.254
C1	0.512	0.575	0.462	0.626	0.648
C2	0.131	0.286	0.090	0.546	**0.001**
C3	0.051	0.448	**0.050**	0.153	**0.002**
C4	<**0.001**	0.371	0.648	0.153	**0.002**
C5	0.055	0.184	< **0.001**	0.463	0.539
C6	0.055	0.184	< **0.001**	0.912	0.115
C7	<**0.001**	0.551	0.390	**0.026**	0.620
C8	**0.001**	0.029	0.488	0.600	0.150
C9	**0.007**	0.078	0.687	0.134	0.074
C10	0.356	0.795	0.120	0.713	0.262
C11	0.585	0.926	0.833	0.781	0.577
C12	0.513	0.367	**0.001**	0.073	**0.033**
C13	<**0.001**	0.732	**0.004**	0.276	0.283
C14	**0.012**	0.711	< **0.001**	0.980	0.375
C15	**0.015**	**0.001**	0.063	0.878	0.842
C16	0.239	0.507	0.714	0.887	0.537
C17	**0.001**	0.073	0.184	0.530	0.683
C18	0.632	**0.003**	< **0.001**	0.408	0.738
C19	**0.008**	0.846	0.125	0.159	0.956
C20	0.524	0.303	0.393	0.557	0.529
C21	0.782	0.349	0.953	0.908	0.435
C22	0.868	0.201	0.328	0.585	0.464
C23	0.087	0.133	**0.001**	0.228	0.398
C24	< **0.001**	0.749	**0.024**	0.266	0.633
Statistics:					
R^2	0.460	0.858	0.722	0.958	0.903
$Adjusted\, R^2$	0.408	0.813	0.632	0.796	0.762
p-value of F-test	< **0.0001**	< **0.0001**	< **0.0001**	**0.0174**	**0.0001**

ular they are: C3, C5, C6, C12, C13, C14, C23, C24.

Observing the main statistics of the linear regression in Table 5.2, it is possible to note that the values of R^2 and $Adjusted\,R^2$ are not very high. This means that the independent variables of co-creation are not able to fully explain the trend of the dependent variable of innovation. On the other hand, the p-value of F-test is very low and this identifies a high statistical significance of the CO model for the FTSE dataset.

5.2.4 DAX

Observing in the Table 5.2 the p-values of the independent variables of the DAX dataset, we note that only one of them is less than 0.05. For this reason, only the variable C7 is statistically significant.

From the statistics of the linear model, it is possible to observe a remarkable difference between R^2 and $Adjusted\,R^2$: this is due to a situation in which the number of observations is too small to estimate the set of parameters. In this case the number of independent variables affects the value of R^2. Also in this case the p-value of F-test assumes a low value, identifying a high statistical significance of the CO model for the DAX dataset.

5.2.5 CAC

Regarding the CAC dataset in Table 5.2 it is possible to notice that some p-values of the independent variables are less than 0.05. Therefore, the related variables are statistically significant and they are: C2, C3, C4, C12.

As in the previous dataset, also in this case the statistics show a significant difference between R^2 and $Adjusted\,R^2$: this means that the number of independent variables in the linear model affects the value of R^2. Also in this case the p-value of F-test assumes a small value and this indicates the presence of a high statistical significance.

The model just described report interesting results regarding the relationship between value co-creation and interest of innovation. In this thesis the same data will be analyzed through the application of Artificial Neural Networks and the results will be compared.

5.3 CO-GE model

In this section we are describing the model in which the predictor set is composed of all co-creation components (C1-C24) plus the absolute number of male (G1) and female (G2) in the board of directors of the business taken into account: this model combines co-creation components and consider gender aspects, and is referred to as CO-GE. This model can be explained as follows:

$$I = \beta_0 + \sum_{i=1}^{k} \beta_i \cdot C_i + \sum_{i=1}^{2} \beta_{k+i} \cdot G_i + \epsilon \tag{5.4}$$

where:
I stands for the articulation of innovation;
β_0 represents the intercept of the regression line;
$k = 24$ is the number of co-creation variables;
β_i is the coefficient related to co-creation component i;
β_{k+i} is the coefficient related to gender component i;
C_i is the value co-creation component i;
G_1 is the coefficient related to the number of male on the board of directors;
G_2 is the coefficient related to the number of female on the board of directors.
ϵ represents an error term.

Table 5.3 shows the coefficients for each of the 24 value co-creation variables and the 2 gender component variables, for all sets of data taken into account, along with the other statistics reported in the previous tables.

In the next subsections we will comment results from Table 5.3, with particular attention to each single set of data.

5.3.1 Eclipse

In this section we will consider the linear model CO-GE regarding the dataset Eclipse.

First of all from Table 5.3 it is possible to observe the p-values related to the value co-creation and gender component variables. Taking into account a 95% confidence interval, the p-values must be less than or equal to 0.05 in order to assert significance of the variables. Table 5.3 shows that this condition is met by only a few variables, namely: C4, C7, C8, C9, C13, C14, C15, C17, C19, and C24. All these variables are related to the aspect of value co-creation and from the observation of the p-values it appears that

Table 5.3: Estimates of the models CO-GE: estimated parameters, p-values and significance statistics of the linear model.

Estimates	Eclipse	NASDAQ	FTSE	DAX	CAC
		Linear model			
Parameters:					
Intercept	-4.1437	1.2634	-1.3754	6.2771	-7.8925
C1	-0.5627	0.1844	-0.3579	2.5212	1.3646
C2	1.1695	0.3175	3.3165	5.3685	17.9499
C3	-1.9233	0.2725	1.3252	2.0419	9.5637
C4	-20.5127	-1.4467	1.2225	2.0419	9.5637
C5	-3.1248	1.4266	2.9868	1.9636	1.6245
C6	-3.1248.	1.4266	2.9868	-0.2354	-6.0580
C7	4.2589	0.1785	-0.4703	2.7491	1.1349
C8	-11.3412	-3.7382	-1.7918	-4.6021	-96.2027
C9	2.8478	1.0147	0.2598	2.9861	-16.1063
C10	-1.1368	0.0560	-2.1767	-0.9866	-5.2591
C11	-1.3556	0.3061	-0.5332	-9.9831	-5.9780
C12	1.0277	-0.6051	-13.8023	51.7845	22.8275
C13	3.2693	0.1618	2.1056	-2.9871	0.9812
C14	-1.2931	-0.2710	3.3577	0.1679	0.8814
C15	4.3344	2.9940	4.4738	-0.9391	-0.8788
C16	6.2534	1.0956	-0.7626	-8.0867	9.0613
C17	-4.9335	-0.8476	-1.6915	-2.1550	1.1952
C18	0.5957	-4.0070	-8.0472	16.3492	-0.3055
C19	-6.3125	0.0658	1.9238	-10.3557	-0.0817
C20	1.7789	2.2388	-1.1176	-35.5952	16.2169
C21	0.5433	1.1092	-0.0397	0.5219	-5.0375
C22	0.4984	1.1765	1.4245	-11.9400	-9.2925
C23	-15.0079	1.7791	21.0236	-53.5615	54.8720
C24	40.0454	0.9462	-7.4456	7.2899	-5.2174
G1	0.1189	-0.4807	0.2291	-1.5937	-0.3976
G2	3.4414	0.5314	-0.8002	0.8200	0.9109
$p-values$:					
Intercept	0.370	0.711	0.844	0.612	0.584
C1	0.512	0.407	0.442	0.537	0.593
C2	0.131	0.318	0.086	0.415	**0.002**
C3	0.051	0.335	**0.050**	0.235	**0.005**
C4	$<$ **0.001**	0.333	0.571	0.235	**0.005**
C5	0.055	0.185	$<$ **0.001**	0.775	0.627
C6	0.055	0.185	$<$ **0.001**	0.974	0.246
C7	$<$ **0.001**	0.503	0.462	0.058	0.573
C8	**0.001**	**0.028**	0.472	0.840	0.188
C9	**0.007**	0.084	0.753	0.438	0.088
C10	0.356	0.874	0.119	0.691	0.263
C11	0.585	0.749	0.829	0.662	0.579
C12	513	0.411	**0.001**	0.107	**0.034**
C13	$<$ **0.001**	0.550	**0.003**	0.325	0.333
C14	**0.012**	0.562	$<$ **0.001**	0.829	0.663
C15	**0.015**	**0.001**	0.053	0.917	0.912
C16	0.239	0.657	0.852	0.750	0.542
C17	**0.001**	0.081	0.174	0.407	0.786
C18	0.632	**0.002**	$<$ **0.001**	0.407	0.964
C19	**0.008**	0.942	0.144	0.254	0.977
C20	0.524	0.292	0.337	0.513	0.674
C21	0.782	0.326	0.940	0.779	0.511
C22	0.868	0.175	0.441	0.525	0.485
C23	0.087	0.182	**0.001**	0.365	0.473
C24	$<$ **0.001**	0.579	**0.019**	0.552	0.694
G1	0.880	0.265	0.767	0.530	0.768
G2	0.081	0.448	0.473	0.906	0.488
Statistics:					
R^2	0.460	0.860	0.725	0.964	0.906
$Adjusted\, R^2$	0.408	0.812	0.625	0.737	0.739
p-value of F-test	$<$ **0.001**	$<$ **0.001**	$<$ **0.001**	0.084	$<$ **0.001**

they are able to affect the aspect related to the articulation of innovation. On the other hand, as far as the gender component is concerned, none of the variables is statistically significant: therefore, none is able to affect the articulation of innovation.

Then Table 5.3 shows the results of some statistical measures. Among these we can observe that R^2 and $Adjusted\, R^2$ assume quite low values. This means that the variables of value co-creation and gender component are not able to explain the trend of the articulation of innovation.
On the other hand, observing the p-value of F-test it can be noticed that it is very small and this indicates that there is a high statistical significance of the linear regression model.

5.3.2 NASDAQ

In NASDAQ dataset, observing the p-values relative to the variables of value co-creation and gender component in Table 5.3 it can be seen that only three of them are less than 0.05: these are relative to variables C8, C15 and C18. Also in this case the variables able to affect the articulation of the innovation are relative only to the aspect of value co-creation. The gender component variables again do not seem to affect the articulation of innovation.

Regarding the other statistical measures, Table 5.3 shows that R^2 and $Adjusted\, R^2$ are quite high, meaning that the variables of value co-creation and gender component are able to explain the trend of the articulation of innovation.
Furthermore, the p-value of F-test is very small, which means that there is a high statistical significance of the linear regression model.

5.3.3 FTSE

With regard to FTSE dataset Table 5.3 shows that also in this case some variables have a p-value lower than 0.05: in particular they are C3, C5, C6, C12, C13, C14, C18, C23 and C24. Once again, the variables that can affect the articulation of innovation are related to value co-creation. Instead, from observing the p-values the variables related to the gender component they are not statistically significant.

Looking at the other statistical measures, it can be seen that R^2 and $Adjusted\, R^2$ are not very high, which means that the trend in articulation of innovation can be partially explained by the value co-creation and gender component.
Also in this case, the p-value of F-test is very small, indicating a high statistical significance of the linear regression model.

5.3.4 DAX

Looking at the p-values in DAX dataset reported in Table 5.3 it is possible to see that none of them is less than 0.05. Therefore, none of the variables related to value co-creation and gender component is statistically significant according to the analysis of p-values. Therefore, it is possible to state that none of the variables affect the the trend of articulation of innovation.

As for the other statistical measures, Table 5.3 shows a significant difference between R^2 and $Adjusted\, R^2$ and this means that the number of independent variables in the linear model affects the value of R^2.
This dataset is the only one that has a p-value of F-test greater than 0.05, meaning that there is a low statistical significance of the linear model.

5.3.5 CAC

Regarding the p-values of CAC dataset, only four of them are smaller than 0.05: they are related to variables C2, C3, C4 and C12. Also in this case the only significant variables according to the analysis of p-values are related to value co-creation: this means that only this aspect is able to affect the articulation of innovation.

Observing the other statistical measures it is possible to notice that there is a remarkable difference between R^2 and $Adjusted\, R^2$: the value of R^2 is affected by the number of independent variables is taken into consideration. Despite the statistical significance witnessed by the p-value of the F-test, our models may be affected by over-fitting, and in order to assess this phenomenon we have computed the $Predicted\, R^2$, which can be defined as a statistical measure that indicates how well a regression model is able to predict responses for new observations. It measures whether the regression model works only with the original data or also with new data. In both CO e CO-GE the $Predicted\, R^2$ values were comprised between 0 and 0.3, which is lower than both R^2 and $Adjusted\, R^2$.

The results just described report interesting information regarding the relationship between the aspects of value co-creation, gender and articulation of innovation. Throughout this thesis these results will be compared to those obtained from the application of Artificial Neural Networks, that will be introduced in the next chapters.

Chapter 6

Our Neural Network Approach

In this chapter we are defining a neural network approach to model the articulation of innovation, by using different combinations of variables as input of the networks. We are using the same sets of predictors identified in Chapter 5, and defining three neural network approaches in which:

- the predictor set is composed of the absolute number of male and female in the board of directors of the business taken into account (defining the *GE* model);
- the model in which the predictor set is composed of all co-creation components (C1-C24) (defining the *CO* model);
- the model in which the predictor set is composed of all co-creation components (C1-C24) plus the absolute number of male and female in the board of directors of the business taken into account: (defining the *CO-GE* model).

We will detail the pre-processing operations needed before to run our algorithms in Section 6.1, before discussing the experimental setting of the network on Section 6.2. The network topologies used in our experimental phase will be outlined in Section 6.3; the partitioning of data between train and validation set will be dealt with in Section 6.4, and the performances of the network will be described in Section 6.5.

6.1 Data Pre-processing

The analysis of data at hand represents an important phase in the experimental settings for a neural network approach: this is done to understand data features, to detect eventual anomalies, and to preserve the most information as possible without getting trapped in over-fitting problems. We have devised, for our approaches, the pre-processing operations defined by di Tollo et al. [41, 42, 40], that we outline in what follows:

Removal and replacement The issue of incurring in missing and wrong values is well-known by all practicioners that deal with real-world applications, and all neural networks approaches proposed by the literature resort to procedures to take into account this aspect. In this thesis, we have used the approach introduced by [41, 42, 40], that suggest to remove indicators containing more than 30% of missing and wrong values. All sets of data introduced in Section 2.2 have been collected for the sake of this thesis and do not show this portion of missing and wrong data, and this means that we are using, in the experimental phase, the whole set of variables.

As for the portion of wrong data, no wrong data has been detected in our sets of data; as for missing data, we want to remark (again) that for some busineses it was not possible to collect all data referring to the gender components, so we have some missing data referring to the number of male and female in the board of directors: in this last case, we have replaced the missing values with the variable's average over all businesses [7, 31].

Normalization When applying neural networks, a widely used rule-of-thumb imposes to perform data-normalization in order to feed the neural network with data belonging to the same range. Many mathematical formulations have been suggested to this aim (see for instance [70]). In our thesis we are using the logarithmic transformation used by di Tollo et al. [41, 42, 40], which is defined as follows: let x_i be the value before normalisation of input x for business i, and $\overline{x}_i$ be its normalised value. The relation between normalised and pre-normalised data can be defined as follows:

$$\overline{x}_i = \log_u (x_i + 1) , \tag{6.1}$$

where $u = x_{\max} + 1$, in order to have $\overline{x}_i \in [0, 1]$. We want to remark that the original formulation proposed by Angelini et al. [7] was the following:

$$\overline{x}_i = \log_u (|\min(0, x_{min})| + x_i + 1) , \tag{6.2}$$

and this was due to the fact that the authors have tackled a problem that used a different variables set, in which there were observations whose values were negative. All values belonging to our sets of data represent weighted occurrences, and they cannot be negative by definition, hence we are not considering the possibility of encountering negative values, that would hinder the application of a logarithmic transformation.

Tables 6.1 – 6.3 shows the main statistics of data at hand after the preprocessing operations.

6.2 Experimental setting

In this section we are defining the Neural Network approach that we are using in the experimental phase. As we have seen in Section 4, Artificial Neural Networks are algorithms whose behavior mimics the behavior of the biologic brain, in order to perform complex tasks. They are generally referred to as black-boxes and they are used to detect non-functional relationships. Neural networks have to perform a *learning* phase, that may be defined according to three different paradigms: *supervised* learning [25]; *unsupervised* learning [13] and *reinforcement* leaning [116].

For our learning phase, we are resorting to *supervised* learning, and this is due to the nature of data at hand: since we aim to predict a given variable (i.e., the articulation of innovation), we can compute exactly the deviation between the output of the Neural Network and the desired output. For this reason, we will be using a neural network approach that resorts to Back-Propagation (see Section 4.3.2).

6.3 Neural Network Topology

As for the Neural Network topology, we have used the *feed-forward* Neural Network, which is still one of the most used architectures, and that has also been used by the existing literature on co-creation and innovation by [42, 40]. In order to define our neural network topology, we had to make some design choices regarding the topology and the learning algorithm, that we are detailling in what follows.

All algorithms design choices can be reconducted to parameter setting problems, whose approaches can be partitioned into three categories: *parameter setting*, *parameter control*, and *fine-tuning*.

In *parameter setting*, the parameters of the algorithm are set before the run, and for this reason these procedures are also referred to as *off-line learning* [66, 47], and this can be done by using several methods. Amongst these models, we recall the simple generate-evaluate methods [66], racing algorithms [19] (that evaluates the performances of a set of parameter candidate solutions, by iteratively removing the ones leading to poor performance),

Table 6.1: Main statistics of the response and explanatory variables after pre-processing operations, for the sets of data Eclipse and NASDAQ.

	Variable name	Average	STD	Min	Max
Eclipse Number of businesses = 287					
	$C1$	0.30	0.24	0	1
	$C2$	0.27	0.25	0	1
	$C3$	0.18	0.18	0	1
	$C4$	0.25	0.25	0	1
	$C5$	0.31	0.27	0	1
	$C6$	0.16	0.25	0	1
	$C7$	0.32	0.24	0	1
	$C8$	0.23	0.24	0	1
	$C9$	0.31	0.31	0	1
	$C10$	0.21	0.20	0	1
	$C11$	0.26	0.26	0	1
	$C12$	0.26	0.27	0	1
	$C13$	0.22	0.20	0	1
	$C14$	0.29	0.24	0	1
	$C15$	0.22	0.24	0	1
	$C16$	0.23	0.24	0	1
	$c17$	0.34	0.28	0	1
	$C18$	0.15	0.19	0	1
	$C19$	0.20	0.24	0	1
	$C20$	0.11	0.18	0	1
	$C21$	0.25	0.25	0	1
	$C22$	0.05	0.11	0	1
	$C23$	0.23	0.24	0	1
	$C24$	0.23	0.25	0	1
	I	0.12	0.20	0	1
	$G1$	0.31	0.29	0	1
	$G2$	0.38	0.29	0	1
	Variable name	Average	STD	Min	Max
NASDAQ Number of businesses = 98					
	$C1$	0.35	0.29	0	1
	$C2$	0.27	0.26	0	1
	$C3$	0.28	0.26	0	1
	$C4$	0.23	0.24	0	1
	$C5$	0.33	0.25	0	1
	$C6$	0.16	0.23	0	1
	$C7$	0.38	0.25	0	1
	$C8$	0.23	0.24	0	1
	$C9$	0.30	0.27	0	1
	$C10$	0.30	0.28	0	1
	$C11$	0.25	0.26	0	1
	$C12$	0.27	0.27	0	1
	$C13$	0.31	0.32	0	1
	$C14$	0.21	0.19	0	1
	$C15$	0.19	0.22	0	1
	$C16$	0.22	0.24	0	1
	$C17$	0.37	0.26	0	1
	$C18$	0.12	0.17	0	1
	$C19$	0.23	0.25	0	1
	$C20$	0.18	0.26	0	1
	$C21$	0.25	0.27	0	1
	$C22$	0.25	0.27	0	1
	$C23$	0.12	0.19	0	1
	$C24$	0.23	0.26	0	1
	I	0.12	0.20	0	1
	$G1$	0.77	0.15	0	1
	$G2$	0.60	0.19	0	1

Table 6.2: Main statistics of the response and explanatory variables after pre-processing operations, for the sets of data FTSE and DAX.

	Variable name	Average	STD	Min	Max
FTSE Number of businesses = 95					
	$C1$	0.48	0.30	0	1
	$C2$	0.30	0.30	0	1
	$C3$	0.24	0.25	0	1
	$C4$	0.24	0.26	0	1
	$C5$	0.33	0.28	0	1
	$C6$	0.13	0.22	0	1
	$C7$	0.48	0.28	0	1
	$C8$	0.19	0.25	0	1
	$C9$	0.24	0.24	0	1
	$C10$	0.33	0.30	0	1
	$C11$	0.26	0.26	0	1
	$C12$	0.19	0.25	0	1
	$C13$	0.30	0.26	0	1
	$C14$	0.32	0.25	0	1
	$C15$	0.24	0.27	0	1
	$C16$	0.19	0.24	0	1
	$C17$	0.34	0.26	0	1
	$C18$	0.18	0.21	0	1
	$C19$	0.25	0.27	0	1
	$C20$	0.14	0.18	0	1
	$C21$	0.20	0.21	0	1
	$C22$	0.13	0.19	0	1
	$C23$	0.18	0.24	0	1
	$C24$	0.18	0.25	0	1
	I	0.10	0.18	0	1
	$G1$	0.74	0.11	0	1
	$G2$	0.72	0.19	0	1
	Variable name	Average	STD	Min	Max
DAX Number of businesses = 30					
	$C1$	0.53	0.23	0	1
	$C2$	0.41	0.24	0.04	1
	$C3$	0.42	0.25	0	1
	$C4$	0.45	0.28	0	1
	$C5$	0.43	0.26	0.01	1
	$C6$	0.33	0.30	0	1
	$C7$	0.64	0.22	0	1
	$C8$	0.36	0.27	0	1
	$C9$	0.35	0.26	0	1
	$C10$	0.56	0.28	0	1
	$C11$	0.42	0.29	0	1
	$C12$	0.38	0.29	0	1
	$C13$	0.51	0.25	0	1
	$C14$	0.51	0.30	0	1
	$C15$	0.47	0.32	0	1
	$C16$	0.37	0.28	0	1
	$C17$	0.48	020	0.02	1
	$C18$	0.37	0.30	0	1
	$C19$	0.40	0.30	0	1
	$C20$	0.37	0.29	0	1
	$C21$	0.47	0.25	0	1
	$C22$	0.22	0.28	0	1
	$C23$	0.37	0.27	0	1
	$C24$	0.38	0.29	0	1
	I	0.15	0.19	0	1
	$G1$	0.79	0.14	0.30	1
	$G2$	0.50	0.29	0	1

Table 6.3: Main statistics of the response and explanatory variables after pre-processing operations, for the sets of data CAC.

	Variable name	Average	STD	Min	Max
CAC Number of businesses = 40					
	$C1$	0.54	0.26	0	1
	$C2$	0.36	0.32	0	1
	$C3$	0.31	0.32	0	1
	$C4$	0.28	0.28	0	1
	$C5$	0.35	0.33	0	1
	$C6$	0.21	0.24	0	1
	$C7$	0.46	0.23	0	1
	$C8$	0.26	0.30	0	1
	$C9$	0.31	0.31	0	1
	$C10$	0.26	0.25	0	1
	$C11$	0.28	0.30	0	1
	$C12$	0.27	0.32	0	1
	$C13$	0.35	0.30	0	1
	14	0.37	0.31	0	1
	$C15$	0.17	0.24	0	1
	$C16$	0.26	0.31	0	1
	$C17$	0.44	0.25	0	1
	$C18$	0.24	0.29	0	1
	$C19$	0.37	0.32	0	1
	$C20$	0.23	0.28	0	1
	$C21$	0.34	0.31	0	1
	$C22$	0.19	0.28	0	1
	$C23$	0.26	0.31	0	1
	$C24$	0.26	0.30	0	1
	I	0.13	0.24	0	1
	$G1$	0.82	0.10	0.53	1
	$G2$	0.72	0.23	0	1

Param-ILS [61] (that performs a local search on the space of parameters). It is out of the scope of this thesis to perform an enumeration of tuning methods, and we forward the interested reader to [47].

In *parameter control*, the parameters of the algorithm are set during the run, and for this reason these procedures are also referred to as *on-line learning*. These procedures encompass both general-purpose methods, along with methods that are designed for a specific algorithms [40]. We forward the interested reader to [4] for a review of parameter control methods.

In *fine-tuning* approaches, the user sets the parameter of the algorithms according to its previous experience and to rules-of-thumbs. Although naïve, this strategy is still used by many practicioners.

As for the topology, the most important parameter to be set is the number of hidden neurons, since the input and output neurons are defined by the inputs and by the desired output of the problem. The literature on the topic has introduced all approaches for parameter setting to this issues, from recognised rules of thumbs to adaptive procedures to change the topology of the neural network over time [44]. We have decided to use a rule of thumb in order to set the number of hidden neurons, as reported by [38] and we have set the number of hidden neurons equal to the number of parameters, by using only one hidden layer. The reason for choosing one layer only is due to the small cardinality of the data at hand, that could hinder the learning algorithm to learn the higher number of parameters associated to the synapses. We know that two-layers neural networks are formally recognised as universal function approximators [63], but the good fit reported in Chapter 5 about linear regression models suggest that one hidden layer could be enough for our purposes.

As for the neural network learning parameter, the literature on the topic reports that the typical values for a neural network with standardized inputs (or inputs mapped to the [0, 1] interval, which is the case in our input definition) has to be smaller than 1 and greater than 10^{-6}[17]; furthermore, there is no way of determining the learning rate a-priori [107]. For these reasons, we have resorted to a parameter tuning procedure to set the learning rate: F-Race, whose execution set the learning parameter to 0.12 for the CO model and to 0.10 for the CO-GE model. Please notice that for both models the momentum value found by F-Race was close to 0, hence we have not introduced any momentum parameter into our analysis. The graphical representation of the two Neural Networks can be found in Figure 6.1 – 6.2.

The Neural Network has been implemented in Python (see Chapter 5), and an outline of the code can be found in Listing 6.1.

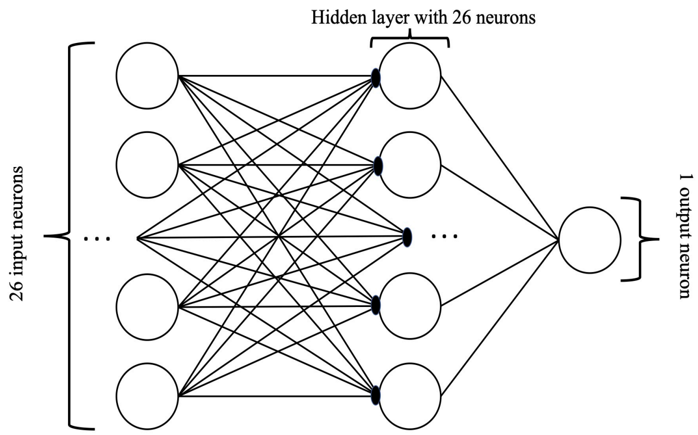

Figure 6.1: The feed-forward Neural Network used for the CO-GE model.

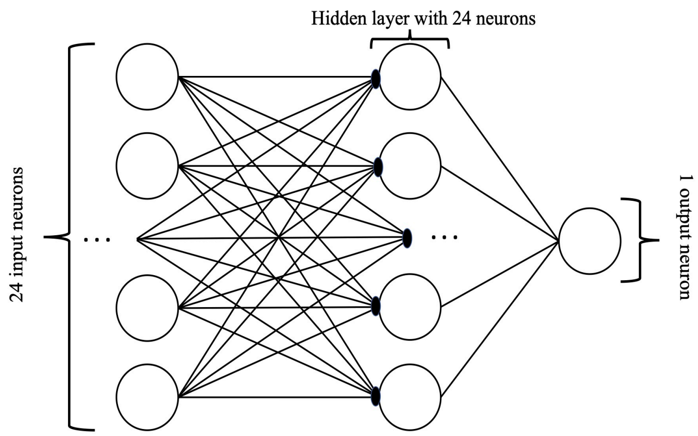

Figure 6.2: The feed-forward Neural Network used for the CO model.

Listing 6.1: Neural Network Python's code

```
import tensorflow as tf
from tensorflow import keras
import dataset
import pandas as pd
import numpy as np
import os

# Do not output verbose messages

os.environ["TF_CPP_MIN_LOG_LEVEL"] = "2"

# Launch of TENSORFLOW

gpus = tf.config.experimental.list_physical_devices('GPU')
if gpus:
    try:
```

```
        tf.config.experimental.set_visible_devices(gpus[0], 'GPU')
        logical_gpus = tf.config.experimental.list_logical_devices('GPU')
        print(len(gpus), "Physical␣GPUs,", len(logical_gpus), "Logical␣GPU")
    except RuntimeError as e:
# Please remember that visible devices must be set before GPUs have been initialized
        print(e)

# Create the folder to store error statistics

if os.path.isdir(r'error'):
    if not os.path.isdir(r'error\dataset0'):
        os.mkdir(r'error\dataset0')
else:
    os.mkdir(r'error')
    os.mkdir(r'error\dataset0')

# Create dataset from loaded data

default_x_data, default_y_data = dataset.get_full_xy_data()

def build_model():

# Neural Network model

    model = keras.models.Sequential()

# Random initialization of network weights, that are generated by using a random uniform distribution from 0 to 1


kernel_initalizer = tf.keras.initializers.RandomUniform(minval=0, maxval=1)

# Definition of Input, Hidden, Output layers;
# Definition of algorithm parameters

model.add(keras.layers.Dense(x_data.shape[1], activation='sigmoid',
input_shape=(x_data.shape[1],),
kernel_initializer=kernel_initalizer))
model.add(keras.layers.Dense(1, activation='sigmoid',
use_bias=False, kernel_initializer=kernel_initalizer))
model.compile(loss=tf.keras.losses.MeanSquaredError(),
optimizer=keras.optimizers.SGD(learning_rate=1e-2))

# print initialised neurons and synapses weights
# print('\nInitialised weights and biases:')
# for layer in model.layers: print(layer.get_weights())

model.summary()
return model
```

6.4 Training and test set

The exploitation of the *supervised* learning paradigm requires that during the *learning* phase we have to provide the network with both the input pattern, and the desired output value. In our case-study, the output pattern consists of the innovation metrics articulation. As for the input pattern, it consists of

- Variables C1-C24 from Tables 6.1 – 6.3, for the experiments that only consider the co-creation components as predictors;
- The number of male and female on the board of directors (G1, G2), along with variables C1-C24 from Tables 6.1 –6.3 for the experiments that consider co-creation keywords and gender components as predictors.

During the neural network learning we have to identify two disjoint sets of observations from overall number of businesses that belong to the sets of data under analysis: the *training set*, that will be used to determine the synapses' weights, and the *validation set*, that is used to determine the network performance and to stop the learning. Several rules have been identified to partition the overall data between these two sets. We have decided to split the overall data by randomly allocating the 70 percent of its observations to the *training set*, and the remaining 30 percent to the *validation set*. This random allocation has been repeated 30 times, each time determining a different *training-validation* partition. The training procedure is outlined in Listing 6.2.

Listing 6.2: Training procedure Python's code

```
# Definition of the training procedure

def train(x, y, index):
    training_error = []
    validation_error = []

# In our experiments we are using 30 different partitions of training-validation sets

# Create the partition, initialize rounds and epochs


for partition_index in range(30):
        test_model = build_model()
        training_x, training_y, validation_x, validation_y =
         dataset.get_training_and_test_set(x, y)
        training_error.append([])
        validation_error.append([])
        print("Partition_" + str(partition_index))
        history = test_model.fit(x=training_x,
                                                            y=training_y, epochs=50, verbose=1, batch_size=8,
                                 validation_data=(validation_x, validation_y), shuffle=False)

# -history- is dictionary that stores the execution trace of every epoch, including the error to be minimised

        training_error[partition_index].extend(np.sqrt(history.history['loss']))
        validation_error[partition_index].extend(np.sqrt(history.history['val_loss']))

# Compute basic statistics

training_error[partition_index].extend(
            [np.min(training_error[partition_index]), np.mean(training_error[partition_index]),
             np.std(training_error[partition_index])])
        validation_error[partition_index].extend(
            [np.min(validation_error[partition_index]), np.mean(validation_error[partition_index]),
             np.std(validation_error[partition_index])])
        columns = ["Round_%i" % x for x in range(50)]
        columns.extend(["Min", "Mean", "Std"])

# Create the output dataframes and save them

        df_train = pd.DataFrame(training_error, columns=columns)
        df_val = pd.DataFrame(validation_error, columns=columns)
        df_train.to_excel
        (r'error\dataset%i\error_train_1t_full_model.xlsx' % index, index=False, header=True)
        df_val.to_excel
        (r'error\dataset%i\error_val_1t_full_model.xlsx' % index, index=False, header=True)

# Pre - Processing Operations

x_data, y_data = dataset.get_full_xy_data()
x_data, y_data = dataset.remove_outliers(x_data, y_data)
x_data, y_data = dataset.replace_missing(x_data, y_data)
x_data, y_data = dataset.normalize(x_data, y_data)

# Train the network

train(x_data, y_data, 0)
```

Table 6.4: Neural Network's overall errors (MSE) on the CO-GE and CO models. We report the statistics on the best runs obtained on the 30 different partitions of overall error.

	CO-GE model				**CO model**			
	Average	STD	Min	Max	Average	STD	Min	Max
Eclipse	0.045	0.004	0.037	0.054	0.045	0.004	0.039	0.054
NASDAQ	0.042	0.006	0.032	0.054	0.043	0.006	0.033	0.053
FTSE	0.035	0.003	0.028	0.040	0.036	0.003	0.030	0.042
DAX	0.045	0.009	0.029	0.068	0.046	0.010	0.030	0.069
CAC	0.114	0.008	0.095	0.129	0.061	0.011	0.045	0.088

```
model.summary()
return model
```

6.5 Algorithm Performances

As for the neural network learning performances, the basic rule consists on operating the supervised learning on the test set, while, in order to avoid overfitting, the goodness of the neural network has to be assessed by computing an error metrics over the validation set, that can be also used in order to set the termination criterion for the algorithm. Several error measures can be used, such as Mean Error (BIAS), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), etc. In our experiments, we have used the network validation set's Mean Square Error (MSE) defined as:

$$\frac{1}{n}\sum_{i=1}^{n}(y-\hat{y}_i)^2, \tag{6.3}$$

where: n is the validation set size, y is acutal output value corresponding to pattern i, and $\hat{y}_i$ is the predicted network output corresponding to input pattern i.

We have then run our neural network approaches on all obtained partitions. For each partition, the algorithm has been run 100 times, and the best run (with respect to the MSE) has been recorded. In order to test the neural network performances and the robustness of the approach, we report in Table 6.4 the statistics of the best runs obtained over the 30 partitions for the experiments run on the CO model and CO-GE models defined in Chapter 5.

Table 6.5: Neural Network's errors (MSE) on the CO-GE model. We report the statistics on the best runs obtained on the 30 different partitions of training-validation set.

	Training set				**Validation set**			
	Average	STD	Min	Max	Average	STD	Min	Max
Eclipse	0.044	0.006	0.036	0.056	0.046	0.012	0.024	0.083
NASDAQ	0.043	0.011	0.022	0.067	0.045	0.019	0.017	0.084
FTSE	0.031	0.009	0.012	0.045	0.036	0.014	0.015	0.078
DAX	0.045	0.019	0.020	0.086	0.059	0.033	0.013	0.110
CAC	0.111	0.029	0.041	0.150	0.121	0.054	0.028	0.233

Table 6.6: Neural Network's errors (MSE) on the CO model. We report the statistics on the best runs obtained on the 30 different partitions of training-validation set.

	Training set				**Validation set**			
	Average	STD	Min	Max	Average	STD	Min	Max
Eclipse	0.047	0.006	0.037	0.059	0.048	0.011	0.026	0.075
NASDAQ	0.040	0.012	0.013	0.058	0.048	0.021	0.014	0.085
FTSE	0.036	0.008	0.018	0.047	0.041	0.028	0.011	0.083
DAX	0.045	0.020	0.007	0.089	0.046	0.034	0.010	0.109
CAC	0.069	0.019	0.028	0.120	0.070	0.034	0.019	0.177

6.6 Discussion of the results

In this chapter we will analyze the results obtained from the application of the neural networks. Tables 6.5 – 6.6 report the main statistics about the MSE obtained by the neural networks, in both training set and validation set. It is possible to state that good results were obtained from the application of the neural networks. The average of the overall errors in Table 6.4 varies between 0.035 and 0.114 in the different datasets for the CO-GE model and between 0.036 and 0.061 in the CO model. In both models, the lowest MSE is found in the FTSE dataset and the highest error is found in the CAC dataset. In addition, looking at the maxima of the overall errors we see that they are not too high: the highest in the CO-GE model is 0.129 for the CAC dataset and 0.088 in the CO model for the same dataset.

These results can be defined as satisfactory with respect to the purposes of this thesis. In fact, our purpose is to understand and classify the different attitudes regarding businesses' interest on innovation. Moreover, the results are satisfactory when compared with the literature in which the relationship between value co-creation and interest of innovation has been studied through the application of neural networks: the MSE is comparable to the

Table 6.7: Overall linear regression's errors (MSE) in CO-GE and CO models for all datasets.

	CO-GE	CO
Eclipse	0.056	0.056
NASDAQ	0.001	0.001
FTSE	0.005	0.005
DAX	0.004	0.004
CAC	0.008	0.008

one found by related literature [39, 42, 40].

From the observation of the Tables 6.5 – 6.6 it is possible to affirm the strength and the ability of generalization of the neural network: their performances have been validated with regards to the validation set, hence on data not encountered during the training of the network. This is made for overcoming the problem of overfitting. We have seen in Chapter 5 that the linear models, used on our sets of data, suffer from this shortcoming, and this is witnessed by low value of the predicted R^2. Furthermore, we observe that the introduction of the gender components into the predictor set leads to better performances in the validation set in three sets of data out of five (Eclipse, NASDAQ, FTSE); on the remaining two (DAX, CAC), its introduction worsen the neural network performances. We can compare this finding with our linear approaches: in that case the introduction of the gender component does not lead to improvements of the R^2 on all sets of data, and in one instance (DAX), the analysis of the $p-value$ of the $F-test$ shows that the regression is not significant when introducing the gender component, whilst when it is not taken into account the regression is always significant. Overall, we can state that the introduction of the gender component does not lead to improve the performances of both linear and neural network approaches.

Results of the neural networks are compared with the analyses performed previously in this thesis. In particular they will be compared to those obtained from the linear regressions described in Chapter 5.
The MSE errors of the neural networks are reported in Table 6.4, while those for the linear regressions are reported in Table 6.7.

In the CO-GE model we observe that the overall average errors are lower in the linear regression (than in the neural networks) results for the NASDAQ, FTSE, DAX and CAC datasets. On the other hand, with regard to

the Eclipse dataset, the value of the lowest overall average error is shown by neural network results. These results confirm what was observed in Chapter 5 when linear regressions were described: the values of R^2 and $Adjusted\,R^2$ were quite high for all datasets except for Eclipse. So the statistical significance of the linear model CO-GE was high for all datasets except Eclipse: the linear model was able to explain for them the relationship between value co-creation, gender component and interest of innovation.
As for the Eclipse dataset, it presented a low value of R^2 and $Adjusted\,R^2$ in the linear model CO-GE. The comparison of the errors between linear regression and neural network shows that with reference to dataset Eclipse the latter is better able to explain the relationship between value co-creation, gender component and interest of innovation.

Also in the CO model, the only dataset in which the overall average error is lower in the neural network results is Eclipse. Also in this model the Eclipse dataset had low R^2 and $Adjusted\,R^2$ values in the linear regression described in Chapter 5. Therefore linear model CO was not able to explain the relationship between value co-creation and interest of innovation with regard to dataset Eclipse. This relationship is better explained by neural network results. On the other hand, in the NASDAQ, FTSE, DAX and CAC datasets the R^2 and $Adjusted\,R^2$ values in the linear regression were quite high: the linear model CO was able to explain the relationship between value co-creation and interest of innovation. From the comparison of the overall average errors it is clear that the linear model is able to better explain this relationship than the neural network with regard to those datasets.

The results just described provide interesting information: depending on the dataset analyzed, the application of a linear model or a neural network is better. In particular, from the overall errors analysis the only dataset that seems to work better with neural networks is Eclipse. Between Eclipse dataset and all the others there is a clear difference in the number of observations: in the first one the number of observations is significantly higher than all the others. This may be a reason why the neural network works better with this dataset. On the other hand, with regard to the introduction of the gender component among the predictors, neural networks seem to give better results than the linear regression model in the Eclipse, NASDAQ and FTSE datasets: in the former the performance of the validation set improves, while in the latter the statistical significance does not increase.

We can conclude that linear models show good fit but also the risk of overfitting. In contrast, neural networks present good errors with regards to the purposes of this thesis, but higher than linear models; however, they are not affected by overfitting. We want to underline that similar works had shown that neural networks had obtained a better fit than linear models, but in these works the application of the two methods was preceded by a phase of application of Principal Components Analysis that we have not introduced in this thesis.

Conclusions

In this thesis, we aimed to model the relationships between gender, value co-creation, and articulation of innovation. In order to achieve this goal, “traditional" methods, i.e., linear regression models, were compared with Artificial Neural Networks.

From a conceptual point of view, first we have decided how to quantify the three aspects of "gender", “innovation", and “co-creation"; then, we have quantified these three aspects for all businesses belonging to all datasets considered; finally, we have used “co-creation" and “gender" as predictors to build a model on "innovation". Regarding the “gender" component, according to the main literature, we have quantified it by counting the number of men and women on the board of directors; regarding "interest in innovation" and "co-creation," we have quantified these aspects by examining the businesses’ communication practices, and by counting the occurrence of a well-defined set of regular expressions. We have quantified the occurrence of these regular expressions in businesses’ websites from different datasets, with the aim of detecting similarities and differences amongst our sets of data. We have selected four sets of data corresponding to stock indices, and one set of data composed of businesses focused on Open-Source development. The choice of the five sets of data was made because: the four sets of data corresponding to the stock indices are able to represent in a general way the economic and entrepreneurial situation of their countries (i.e., USA, UK, Germany and France); the set of data composed of businesses focused on Open-Source development represents a list of companies that are strongly innovation-driven. One of the objectives of this thesis was to understand if as a result of the experimental analysis we could find differences between the different sets of data, with particular attention to the difference between the businesses focused on Open-Source and those belonging to the stock indices. Before beginning the actual experimental phase, key statistics (i.e., Average, Standard Deviation, Minimum, Maximum) were calculated for each aspect of co-creation, gender, and articulation of innovation in each of the five sets of data. In addition, initial correlation analyses were conducted between 1) the ratio of women to total in the board of directors and articulation of innovation, 2) the number of men in the board of directors and articulation of innovation, 3) the number of women in the board of directors and articulation of innovation. From a preliminary analysis it was not possible to draw relevant conclusions on the relationship between gender component, co-creation component and articulation of innovation, but some differences emerged between the five sets of data. Correlation matrices were then constructed considering all the variables. We have used linear regression and neural networks to model the relationship between gender, articulation of innovation, and co-creation. Three models were constructed and analyzed that included 1) the gender component and articulation of innovation, 2) the co-creation component and articulation of innovation, and 3) the gender component, co-creation component, and articulation of innovation. Finally, we compared the results obtained from these two analyses.

Before to draw our final conclusions, some important takeaways have to be taken into account:

First, as in all "analytics" procedure, the application of computational methods to real-world scenarios have to devote a big attention to the data collection part, followed by a great emphasis on pre-processing operations.

Second, the behavior of computational methods have to take into account several instances (benchmarks) in order to provide useful insights

Third, the application of Neural Networks has to coupled with a good understanding of the state-of-the art and with a knowledge of the problem at hand: they are black-boxes and their results can be fuzzy, or anyhow difficult to interpret, and they have to be compared with all available information.

After the experimental phase, it is not possible to clearly detect one method that outperforms the other: depending on the dataset analyzed, the application of a linear model or a neural network is better. Regarding the relationship between the aspects of "co-creation" and "innovation" we have found some interesting results, which confirm that this relationship exists over different benchmarks. Regarding the relationship between the aspect of "gender" and "innovation" it is not possible to draw the same conclusion, hence this aspect will be further investigated in further works: overall, we can state that the introduction of the gender component does not lead to significantly improve the performances of both linear and neural network approaches; moreover, the relationship between gender component and articulation of innovation alone does not appear to be significant. Linear models show significant performances but can suffer over-fitting; neural networks offer results that are comparable with those obtained by the related literature and are not affected by this shortcoming. It is anyhow interesting to remark that the best performances over the Eclipse set of data, which is composed of businesses highly associated with innovation, are shown by Neural Networks. This may confirm that the relationship between gender, co-creation and innovation is nonlinear on innovative businesses.

Appendix

The present appendix contains some methods and models that were useful for the purpose of writing this thesis.

Correlation

During the work for my thesis, I performed correlation analyses using Pearson and Rank Based measures. In what follows I will introduce them and provide some insight about their use.

Correlation is "the state or relation of being correlated", or, more specifically "a relation existing between phenomena, things, between mathematical or statistical variables: they can vary, be associated, and occur together in a way not expected on the basis of chance alone"[9]. It can be asserted that correlation is a statistical measure that explains the relationship between two variables, by defining the rule according to which when one variable changes, the other changes as well.

The correlation measure is also called *correlation coefficient*; the correlation coefficient related to two variables x and y is denoted by $\rho_{x,y}$ and is calculated as follows:

$$\rho_{x,y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(N-1)\sigma_x\sigma_y} = \frac{\sigma_{x,y}}{\sigma_x\sigma_y} \tag{6.4}$$

where: σ_x is the standard deviation of x; σ_y is the standard deviation of y; $\sigma_{x,y}$ is the covariance between x and y.

This measure is called *Pearson's product moment correlation.* Pearson's correlation can calculate the linear relationship between the two variables x and y. In addition, for this measure to give valid results the two variables

[9]These correlation definitions are given by Merriam-Webster Dictionary (https://www.merriam-webster.com/dictionary), accessed on January 21^{st}, 2021.

must be normally distributed. *Pearson's product moment correlation* can assume values between -1 and $+1$, and its value can be interpreted as follows:

- If it assumes value -1 it means that there is an inverse relationship between the two variables and this represents the case of maximum discordance. This type of relationship between the two variables can be represented graphically in a Cartesian plane: variable x is on x-axis; variable y is on y-axis; the points in the graph related to the observations of the two variables can be connected by a decreasing line, for which the smallest (biggest) value of a variable corresponds to the biggest (smallest) value of the other one.
- If it assumes value 0 there is an absence of correlation and the possible relationship between the two variables is not explained by this statistical measure. This type of relationship between the two variables can be represented graphically in a Cartesian plane: variable x is on x-axis; variable y is on y-axis; the points in the graph related to the observations of the two variables are distributed randomly.
- If it assumes value 1 there is a direct relationship between the two variables and this represents the case of maximum concordance. This type of relationship between the two variables can be represented graphically in a Cartesian plane: variable x is on x-axis; variable y is on y-axis; the points in the graph related to the observations of the two variables can be connected by a increasing line, for which to the smallest (biggest) value of a variable corresponds the smallest (biggest) value of the other one.

Of course, it is also possible to obtain intermediate values, so if the coefficient is close to the value $+1$ there is a good concordance, while if it is close to -1 there is a good discordance; if it is close to 0 it is not sufficient to explain the correlation between two variables.

Another type of correlation is the *Spearman's rank correlation* [115]: this is a statistical measure which is denoted by r_s. While the Pearson's correlation measure the magnitude of the linear relationship between two variables, the Spearman's correlation coefficient measure the magnitude of the monotonic relationship between two variables. Unlike Pearson's correlation, in Spearman's correlation the variables considered do not need to be normally distributed and for this reason it is a non-parametric statistic.

In order to compute Spearman's rank correlation, the two variables under consideration must be ranked: the highest value in the rank of each variable is given the “first place" and the lowest value in the rank of each variable is given the “last place". Then, the correlation coefficient is calculated on the basis of the ranking position of the value assumed by the two variables for each observation. The same formula as Pearson's correlation is used to calculate Spearman's rank correlation: in this case the formula is not applied to the variables' values, but to the ranks of the variables; so it is applied to ordinal quantitative variables.

Also in Spearman's rank correlation the correlation coefficient can assume values from -1 to $+1$: the value -1 indicates a perfect negative monotonic correlation; the value 0 indicates absence of monotonic correlation; the value $+1$ indicates a perfect positive monotonic correlation. Regarding the intermediate values between -1 and $+1$ different degrees of monotonic correlation can be identified: values close to zero indicate a weak monotonic correlation that can be negative or positive depending on the sign; values close to -1 indicate a strong negative monotonic correlation; values close to $+1$ indicate a strong positive monotonic correlation..

Correlation is widely used because it is relatively easy and quick to apply to identify simple relationships between variables. Therefore, it is very useful when a specific and exclusive cause-effect relationship is not sought in the components being investigated. This is also the major limitation of this statistical measure: it indicates whether there is a relationship between the two variables taken into account, but does not indicate whether this relationship is totally explained by these two, or is also influenced by other components that have not been taken into account for a specific analysis.

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a generative probabilistic model of a corpus, which is defined as a set of documents, and it represents one of the most popular topic modelling methods. It was proposed by Pritchard and Stephens [105] in 2000 and applied to machine learning by Blei, Andrew Ng and Jordan [20] in 2003.

To understand the basic behavior of LDA it is important to analyze the relationships between documents, topics and words. On one hand, every document can be thought of as a probability distribution of topics; every topic, on the other hand, can be thought of as a probability distribution of

words. Starting from a collection of documents, the goal of LDA is to learn the topics structure in each document and the words structure in each topic. The LDA process uses the Dirichlet distribution[10], which is a continuous probability distribution and depends on a vector of positive real numbers α: for this reason it is usually indicated with $Dir(\alpha)$.
In M documents it is possible to identify a number K of topics to which a number N of words are linked: the aim of LDA is to understand the relationships between M , N, and K.

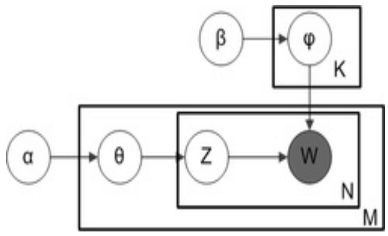

Figure 6.3: Latent Dirichlet Allocation Blueprint.
Source: Anusha Ashok, and Sanjay Singh. “Is that twitter hashtag worth reading." 2015.

In Figure 6.3 it is possible to observe the mechanism thanks to which the Latent Dirichlet Allocation works. The parameter α represents a Dirichlet distribution and, in particular, it refers to the distribution of the various documents in relation to the different topics taken into consideration. β also represents a Dirichlet distribution, but it is the distribution of the various topics in relation to the words in the documents.

Starting from α the process determines a multinomial distribution from which the topics are extracted step by step: this multinomial distribution is referred to as θ. Starting from β, instead, the process obtains another multinomial distribution from which are extracted the words which, in relation to the topics, will form the output of the process: this multinomial distribution is referred to as ϕ. Eventually, the list of topics z is computed from θ, and the combination of z and ϕ represents a list of words, each related to a topic.

At the beginning of the process the user has to set the number of topics (K), and the words in the documents are randomly assigned to the different topics.

The algorithm analyzes each word and the topic it belongs to in each

[10]This process is named after the German mathematician Peter Gustav Lejeune Dirichlet (1805-1859).

document. It assign a word to a topic by taking into account two factors: how often the topic occurs within the document and how often the word occurs within the topic in all documents. It is possible to state, therefore, that during this process two properties are followed. The former consists in the fact that the obtained documents refer to the smallest number of possible topics. The latter maximize the probability that the same word, repeated several times in the documents, is associated with the same topic.

At the end the output shows the number of topics initially chosen by the user. Each topic is associated with the most recurrent words referred to it, together with a probability of belonging of the word to that topic. At this step the user's interpretative ability is fundamental: it is necessary to recognize if each set of words can be rationally led back to a specific topic and, if so, to establish which one. The pseudocode of LDA is outlined in the following Listing 6.3:

Listing 6.3: LDA pseudocode

```
for all topics $k$ in $[1,K]$ do
                sample mixture components $k ~ Dir(\beta)$

for all documents $m$ in $[1,M]$ do

        sample mixture proportion $m ~ Dir(\alpha)$
        sample document length $N_m ~ Poiss(\xi)$
        for all words $n$ in $[1,N_m]$ do

                sample topic index $z_{m,n} ~ Mult(\theta)$
                sample term for word $w_{m,n} ~ Mult(\phi_k)$
```

where:
$Dir(\alpha)$ and $Dir(\beta)$ stand for Dirichlet distribution;
$Poiss(\xi)$ stands for Poisson distribution, which is a discrete probability distribution: in a given time interval, it expresses the probability of the number of events that occur successively and independently;
$Mult(\theta)$ and $Mult(\phi)$ stand for Multinomial distribution, which is a discrete probability distribution: it expresses the probability that events with more than two possible outcomes will occur.

SISEM

SISEM stands for "Self Implication Strategies for Ethics in Management" and it is a model which identifies six motivation sources for individuals (https://www.sisem-institut.com access on 14/04/2021). This model deals with the motivation sources that stimulate human beings during their lives. From observing individuals, it was found that these six motivation sources combined represent the drivers that lead individuals to achieve their goals. SISEM model is able to reveal the motivational profile of an individual: each motivation source corresponds to a specific type of individual, with personal interests well delineated. From these information it is possible to deduct the conditions that stimulate that individual to achieve his/her personal and professional development and success.

The motivation sources belong to three main components of the model: the relational one (“I comuunicate"), the cognitive one (“I think") and the operational one (“I do").

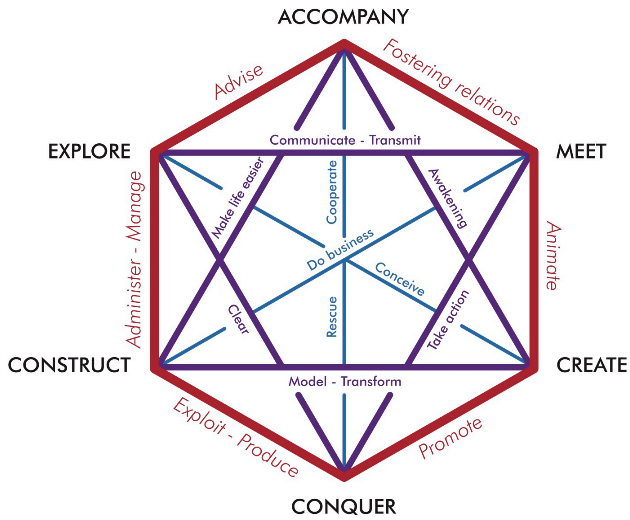

Figure 6.4: The SISEM framework.
Source: https://sisem-institut.com/

The relational component contains two motivation sources: *Accompany* and *Meet*.
An individual driven by the Accompany motivation source likes to relate to others, to share moments of life with them and to offer his/her help in case of need. For this individual it is essential to feel useful for others and human relationships are perceived as mutually beneficial. At work, these people are encouraged to give their best by working in a team to achieve a collective goal.
People driven by the Meet motivation source think that getting to know new and different people is a key resource for their personal and professional development. Diversity in human beings is a source of richness and this makes

each person special and interesting. Such personal profile is stimulated to work better and achieve its goals if put in conditions where it is possible to establish connections with other people, exchange ideas and find new ways of thinking or acting.

The cognitive component contains two motivation sources: *Explore* and *Create*.
Those who are driven by the Explore motivation source love to learn and discover things that are always new and different, and believe that the world can offer them a wide variety of experiences to enjoy and places to visit. They like to investigate, research and verify information and are inclined to analyze everything down to the smallest detail. They are challenged to achieve their goals through the perpetual search for answers (the “why" and the “how") and the resolution of problems that are encountered during their work or personal lives. In addition, they are bored in situations where their own actions become a routine: they need ever-changing tasks to perform.
Individuals driven by the Create motivation source think that the world is not something fixed, but something that can change all the time. At the heart of this thinking there is the power of the imagination, which is a fundamental element for such individuals: they are attracted to the world of ideas, love novelty and take an original approach to every activity they undertake. Moreover, they adapt easily to continuous evolution and wish to be part of it through the discovery and development of innovative things. They are stimulated to work more efficiently in situations where the same things are not done all the time in a repetitive manner, but rather where things are always done in a different way and each day represents a discovery.

Finally, the relational component contains the last two motivation sources: *Build* and *Conquer*.
An individual driven by the Build motivation source likes to be in control of the reality around it. He/she believes that the world can be shaped and controlled in the best way to achieve its goals. This individual prefers to work in situations where his field of action is well defined: everything must be in its place and everyone must know what to do. In order to achieve concrete results he produces an outline and follows a plan, controlling every step of it. In his/her work, this person always tries to aim for efficiency and follow precise rules: everything that involves a certain stability is appreciated by this individual.
Individuals driven by the Conquer motivation source tend to make decisions and subsequent actions very quickly: they often choose to take shortcuts to save time and do not act according to a precise and previously established

plan, but rather make their own decisions on what to do during the process. They need to live in an ever-changing reality, full of action, opportunities to take their own risks and experience the thrills. What motivates them most are situations where goals seem impossible to achieve and no one has ever been able to do so: they are always driven by challenges that include going beyond, doing the impossible and setting new records. They are willing to take any risk in order to achieve such desired results.

The SISEM model is used to analyze the motivational profiles of individuals in a business: the model identifies the motivation sources related to each individual and consequently it is possible to analyze the ties that exist in the group. Through the study of interpersonal ties it is possible to understand the strengths and causes of conflict in the group: individuals guided by certain motivation sources tend to establish a constructive relationship with other individuals guided by certain motivation sources; on the other hand it is likely that conflicts are triggered between individuals guided by other specific motivation sources. For example, it may happen that two people complement each other, so it would be beneficial to have them work together. In an opposite situation, it is possible that two colleagues who are always in close professional contact are motivated by the same components of the model: in this case a conflict may arise that does not lead to good results, or that leads to efficiency anyway; finally, it is possible that in such a situation there is no conflict, but rather a synergy between the two individuals.

Figure 6.4 shows the motivation sources a business has to focus on to achieve specific goals.

For example, if the goal is to introduce *innovation*, it is necessary to focus on the *Explore* and *Create* spheres as far as the cognitive component is concerned. In order to innovate, it is necessary to explore new realities, what does not yet exist, or what already exists but can be applied in a new and revolutionary way within the business. Consequently, it is necessary to create something new in the first case or a new application in the second case. Subsequently, as far as the operational part is concerned, the focus is on the *Conquer* motivation source: it is not possible to follow a pre-established and precise plan in all its steps, since being something new and innovative it is not possible to foresee every detail that may occur during the process, but on the contrary it is necessary to be ready to face in a proactive way every surprise or error that may be found.

If the goal is to practice *co-creation* activities there are several possible patterns to follow.

One possible pattern takes into consideration the *Accompany* motivation source for the relational component, *Explore* and *Create* for the cognitive component. The Accompany motivation sources is recurrent in all the possible patterns since it is fundamental in co-creation activities: during co-creation activities a meeting (physical or virtual) between the business and customers is needed because they plan together ideas dealing with the production of final products or services. In this pattern there are the Explore and Create motivation sources because the business and customers explore the world of ideas and select some of them in order to create a new product that meets the needs of the segment of consumers they want to serve.
In the second pattern there are the *Accompany* motivation source for the relational component, *Create* for the cognitive one and *Conquer* for the operational one. The first two have already been described in the preceding pattern. The third could be traced back to what has been said about innovation: during co-creation activities a new product or service is designed, so it is not possible to be sure of everything that could happen during the process, of what can lead to success and what can be wrong; it is therefore necessary to take the risk of some actions, being aware of everything, ready to give one's best and accept possible failures.
In the last pattern a different combination of the motivation sources is proposed: there are *Accompany* one for the relational component, the *Explore* one for the cognitive component and the *Conquer* one for the operational component.

In addition to innovation and co-creation, other important aspects related to the business activity can be linked to the six motivation sources: these aspects can be related to some value co-creation keywords [40].
The first aspect is *mutual learning* [74] (related to the keywords C5, C12, C21, C24 in Table 2.1), : mutual learning is a key element in the co-creation process because during the collaboration the business learns what the customers' needs are and the customers become aware of the business' products and services. Mutual learning is identified in the *Accompany* motivation source for the relational component, *Explore* for the cognitive one and *Build* for the operational one. In order to achieve mutual learning, it is necessary to get in touch with other people and help them feel comfortable and trusting, making them able to express themselves at their best. Moreover, learning always involves exploring new places, new cultures and new points of view and thanks to this it is possible to build a new knowledge by expanding one's horizons.

The second aspect concerns *resources and processes* [99] (related to the keywords C3, C6 in Table 2.1): in this aspect there are the *Accompany* motivation source for what concerns the relational component and *Build* for the operational one. When talking about human resources the relational component is relevant: human resources must be helped to integrate and interact with each other in order to be able to establish professional ties. This then leads to the construction of processes that are necessary for the functioning of the business itself.

Another important aspect for businesses is *customer relationships* [99] (related to the keywords C7, C11, C20 in Table 2.1). In this case the motivation sources involved are *Accompany* and *Meet* for the relational component, *Explore* and *Create* for the cognitive one and *Build* for the operational one. In order to obtain solid relationships with customers, a meeting opportunity between businesses and customers is needed and it is possible through several channels. The meeting is useful to inform customers about products' or services' characteristics, to enable customers to buy products and receive them, and to be able to give and receive feedbacks. Each business can choose a different level of intensity in the customer relationship, by deciding how much it wants to accompany its customers in their own choices.

Principal Components Analysis

The Principal Components Analysis (PCA) is a statistical technique that was invented in 1901 by Karl Pearson [102] and later developed by Harold Hotelling in 1933 [64] and which is used to reduce in terms of size a dataset taken into consideration when it is considered necessary. This method is commonly used in data science and machine learning, but also in some fields such as Biology, Chemistry, Medicine and many others.

It is often the case that a phenomenon is described by many variables, but that some of them may be redundant or lack the power to influence that phenomenon. PCA aims to make the analyzed data easier to read and to do this it compresses a large number of data into something that is able to capture the essence of the original: through it are taken datasets with many dimensions which are reduced to two or three dimensions. This reduction in dimensionality is done in a meaningful way by grouping the variables according to their similarities and differences and identifying two “new" axes with new directions based on the different degrees of variation found in the data: these new axes represent the *Principal Components* (PCs).

The PCA process begins with the formation of a matrix A in which the rows contain the n observations and the columns contain the k variables under consideration.
The first step is to calculate the mean for each column of A and subtract that mean from the matrix A:

$$B = A - \bar{A} \tag{6.5}$$

Subsequently, it is necessary to calculate the covariance[11] matrix C as follows:

$$C = B^T \times B \tag{6.6}$$

where B^T is the transposed matrix of B.

It is possible to observe that the matrix C has size $k \times k$ and on the main diagonal there are the variances of the k variables.
If the covariance between two variables is positive it means that as one variable increases (decreases) the other one also increases (decreases); on the contrary, if the covariance is negative it means that as one variable increases (decreases) the other one decreases (increases).

The next step is to calculate the eigenvalues and eigenvectors of matrix C.
A value λ is defined as an eigenvalue of the matrix C if:

$$Cv = \lambda v \tag{6.7}$$

where: C is a squared matrix; v is the eigenvector of the matrix C.

Then the eigenvalues of the matrix C are computed using the following formula:

$$det(C - \lambda I) = 0 \tag{6.8}$$

where: det indicates the determinant of a matrix; I is the identity matrix.

The eigenvector matrix v is then also calculated by following the formula below:

$$det(C - \lambda I)v = 0 \tag{6.9}$$

[11] The covariance expresses the variance between two variables.

This matrix v also has size $k \times k$ and each column corresponds to the eigenvector relative to one of the k variables, to which an eigenvalue λ_k can be associated.

At this point the eigenvalues are ranked from the highest to the lowest value: the eigenvectors corresponding to the highest eigenvalues are those that contain more information regarding the distribution of the data. On the contrary, the eigenvectors corresponding to the lowest eigenvalues contain less information and therefore can be eliminated: the number of eliminations depends on the number of dimensionalities that are desired, considering that each eigenvector corresponds to a dimensionality.

Once selected the eigenvectors to be eliminated, the matrix W is constructed with the remaining ones in the following way: in the first column there is the column of the matrix v corresponding to the highest eigenvalue λ_k, and in the following ones those corresponding to the other remaining eigenvalues in descending order of size.
The matrix has dimension $k \times d$, where k is the number of original variables and d is the number of new dimensions to be obtained.

In order to calculate the *Principal Components (PCs)* it is necessary to derive W^T, that is the transposed matrix of W and has as dimension $d \times k$. The PCs are calculated in the following way:

$$PC = W^T \times B^T \tag{6.10}$$

The rows of the PC matrix represent the two new axes in which the original observations are graphically inserted into the subspace reduced in size from the starting one.

These *Principal Components* have some properties:

- They are linear combinations of the original variables, where the weights are represented by the eigenvectors;

- They are orthogonal between them;

- The variation between them decreases as one moves from the first component to the last one.

To conclude it is possible to assert that this process inevitably causes a small loss of information, which, however, is not excessive if the highest eigenvalues to be held are chosen wisely and correspond to the variables that most influence the observations.

Acknowledgements

First of all, I wish to acknowledge all lecturers of the master's degree program in Global Development and Entrepreneurship for their valuable teachings. Then, I wish to acknowledge the Erasmus+ program for giving me the opportunity to do an internship in Belgium and Luxembourg, during which I had the pleasure of meeting some excellent people whom I would like to thank now: Thierry Delperdange, founder of Thierry Delperdange Coaching, where I did my internship; Patrick Apostolo, Corporate Operations Director EMEA at Radisson Hotel Group; Dr. Chiara Orlandi, Research Associate at Institute of Human Virology (University of Maryland); Dr. Joseph Andria, Researcher at University of Palermo.
I would like to thank Davide Marchesan who through the VERA project helped me in the data collection of this thesis.

Then I would like to thank my supervisor, Professor Giacomo di Tollo who showed his belief and trust in me from the very first moment by giving me important tasks and including me in his research work. Thanks to him I was able to do my internship abroad and collaborate with great people. During the writing of my thesis he was always available, he always supported me and allowed me to learn new valuable things. For these reasons I am grateful to him and I hope to continue to collaborate with him.

I am thankful to my parents for going along with my choices by staying by my side and showing their pride in me; to my brother because we have a unique relationship where we show each other everything in our own way and that is my pillar. I am thankful also to my boyfriend who has been fundamental in these two years of my master's degree; he supported me in my studied, added happiness into my life with our adventures, our travels, and our laughter at silly things and understood me in my difficult moments.

I would like to thank my lifelong friends in my beloved Pisa who, despite the distance, are always there; they have always filled my life with unforget-

table moments and every time I move they leave me with the desire to return to meet them.
Then I would like to thank all my friends from all over Italy and the world that I was lucky to meet during my Erasmus period.
Finally, I would also like to thank my classmates with whom studying was more fun; my roommates in Treviso with whom, despite the situation, I had some magical moments.

In the end, I would like to dedicate this work to my grandmother who would have been proud of me.

Bibliography

[1] Andrea E. Abele. The dynamics of masculine-agentic and feminine-communal traits: findings from a prospective study. *Journal of personality and social psychology*, 85(4):768, 2003.

[2] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

[3] George A. Akerlof and Rachel E. Kranton. *Identity economics: How our identities shape our work, wages, and well-being*. Princeton University Press, 2010.

[4] Aldeida Aleti and Irene Moser. A systematic literature review of adaptive parameter control methods for evolutionary algorithms. *ACM Computing Surveys (CSUR)*, 49(3):1–35, 2016.

[5] Deborah Alexander and Kristi Andersen. Gender as a factor in the attribution of leadership traits. *Political Research Quarterly*, 46(3):527–545, 1993.

[6] Stephen Allen, Stoyan Tanev, and Tony Bailetti. Towards the development of a research methodology for studying the nature of value co-creation in internet driven businesses. 2009.

[7] Eliana Angelini, Giacomo di Tollo, and Andrea Roli. A neural network approach for credit risk evaluation. *The quarterly review of economics and finance*, 48(4):733–755, 2008.

[8] David Antons, Eduard Grünwald, Patrick Cichy, and Torsten O. Salge. The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management*, 50(3):329–351, 2020.

[9] Bill Aulet. *Disciplined entrepreneurship: 24 steps to a successful startup*. John Wiley & Sons, 2013.

[10] Charles Baden-Fuller. Strategic innovation, corporate entrepreneurship and matching outside-in to inside-out approaches to strategy research 1. *British Journal of Management*, 6:S3–S16, 1995.

[11] David Ballantyne. Dialogue and its role in the development of relationship specific knowledge. *Journal of Business & Industrial Marketing*, 2004.

[12] Anahita Baregheh, Jennifer Rowley, and Sally Sambrook. Towards a multidisciplinary definition of innovation. *Management decision*, 2009.

[13] Horace B. Barlow. Unsupervised learning. *Neural computation*, 1(3):295–311, 1989.

[14] Antonella Basso and Giacomo di Tollo. Prediction of uk research excellence framework assessment by the departmental h-index. *European Journal of Operational Research*, 2021.

[15] David N. Beede, Tiffany A. Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E. Doms. Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11), 2011.

[16] Bari L. Bendell, Diane M. Sullivan, and Kathrin J. Hanek. Gender, technology and decision-making: insights from an experimental conjoint analysis. *International Journal of Entrepreneurial Behavior & Research*, 2020.

[17] Yoshua Bengio. *Practical Recommendations for Gradient-Based Training of Deep Architectures*, pages 437–478. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[18] John Bessant, Richard Lamming, Hannah Noke, and Wendy Phillips. Managing innovation beyond the steady state. *Technovation*, 25(12):1366–1376, 2005.

[19] Mauro Birattari, Thomas Stützle, Luis Paquete, Klaus Varrentrapp, et al. A racing algorithm for configuring metaheuristics. In *Gecco*, volume 2, 2002.

[20] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of machine Learning research*, 3:993–1022, 2003.

[21] Niels Bosma, Stephen Hill, Aileen Ionescu-Somers, Donna Kelley, Maribel Guerrero, and Thomas Schott. Global entrepreneurship monitor: 2020/2021 global report. Technical report, Global entrepreneurship monitor, 2021. Accessed 15 May 2021.

[22] Leonardo Buzzavo. Strategy innovation as business model reconfiguration. *Department of Management, Università Ca'Foscari Venezia Working Paper*, (2), 2012.

[23] David Card and A. Abigail Payne. High school choices and the gender gap in stem. *Economic Inquiry*, 59(1):9–28, 2021.

[24] Priyanka B. Carr and Claude M. Steele. Stereotype threat affects financial decision making. *Psychological Science*, 21(10):1411–1416, 2010.

[25] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, 2006.

[26] Selcuk Cebi. Determining importance degrees of website design parameters based on interactions and types of websites. *Decision Support Systems*, 54(2):1030–1043, 2013.

[27] Jie Chen, Woon S. Leung, and Kevin P. Evans. Female board representation, corporate innovation and firm performance. *Journal of Empirical Finance*, 48:236–254, 2018.

[28] David Choi and Liisa Valikangas. Patterns of strategy innovation. *European Management Journal*, 19(4):424–429, 2001.

[29] David Choi and Liisa Valikangas. Patterns of strategy innovation. *European Management Journal*, 19(4):424–429, 2001.

[30] Rae Cooper. The gender gap in union leadership in australia: A qualitative study. *Journal of Industrial Relations*, 54(2):131–146, 2012.

[31] Marco Corazza, Davide De March, and Giacomo di Tollo. Design of adaptive elman networks for credit risk assessment. *Quantitative Finance*, 21(2):323–340, 2021.

[32] Catherine M. Daily and Dan R. Dalton. Women in the boardroom: A business imperative. *Journal of Business strategy*, 2003.

[33] Fariborz Damanpour. Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of management journal*, 34(3):555–590, 1991.

[34] Kirsti Dautzenberg. Gender differences of business owners in technology-based firms. *International Journal of Gender and Entrepreneurship*, 2012.

[35] Flavio De Jesus Matias. Php web scraper for regular expressions. Technical report.

[36] Davide De March, Giacomo di Tollo, and Stoyan Tanev. Using online innovation metrics to cluster co-creative firms by soms. In *ISPIM Conference Proceedings*, page 1. The International Society for Professional Innovation Management (ISPIM), 2012.

[37] Vijay S. Desai and Rakesh Bharati. A comparison of linear regression and neural network methods for predicting excess returns on large stocks. *Annals of Operations Research*, 78:127–163, 1998.

[38] Giacomo di Tollo. Reti neurali e rischio di credito: stato dell'arte e analisi sperimentale, 2005.

[39] Giacomo di Tollo and Stoyan Tanev. A neural network approach for the evaluation of the innovation outcomes of value co-creation practices in technology-driven firms. 2010.

[40] Giacomo di Tollo, Stoyan Tanev, Giacomo Liotta, and Davide De March. Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms. *Computers in Industry*, 74:16–28, 2015.

[41] Giacomo di Tollo, Stoyan Tanev, Zheng Ma, et al. Neural networks to model the innovativeness perception of co-creative firms. *Expert Systems with Applications*, 39(16):12719–12726, 2012.

[42] Giacomo di Tollo, Stoyan Tanev, Kassis M. Slim, and Davide De March. Determining the relationship between co-creation and innovation by neural networks. In *Complexity in Economics: Cutting Edge Research*, pages 49–62. Springer, 2014.

[43] Dimiter Dobrev. A definition of artificial intelligence. *arXiv*, 2012.

[44] David Dominguez, Kostadin Koroutchev, Eduardo Serrano, and Francisco B Rodriguez. Dynamical neural network: Information and topology. *arXiv preprint cs/0506078*, 2005.

[45] Peter Drucker. The practice of management. *New York: Harpers*, 1954.

[46] Catherine C. Eckel and Philip J. Grossman. Chapter 113 men, women and risk aversion: Experimental evidence. volume 1 of *Handbook of Experimental Economics Results*, pages 1061–1073. Elsevier, 2008.

[47] Agoston E. Eiben and Selmar K. Smit. Parameter tuning for configuring and analyzing evolutionary algorithms. *Swarm and Evolutionary Computation*, 1(1):19–31, 2011.

[48] Naomi Ellemers. Gender stereotypes. *Annual review of psychology*, 69:275–298, 2018.

[49] Jeffrey L. Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.

[50] Michael Etgar. A descriptive model of the consumer co-production process. *Journal of the academy of marketing science*, 36(1):97–108, 2008.

[51] John E. Ettlie and Ernesto M. Reza. Organizational integration and process innovation. *Academy of management journal*, 35(4):795–827, 1992.

[52] Walter J. Ferrier. Navigating the competitive landscape: The drivers and consequences of competitive aggressiveness. *Academy of management journal*, 44(4):858–877, 2001.

[53] Dario Floreano and Stefano Nolfi. Reti neurali: algoritmi di apprendimento, ambiente di apprendimento, architettura. *Università di Trieste, Giornale italiano di psicologia, XX febbraio*, pages 15–50, 1993.

[54] Carl F. Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss*. sumtibus Frid. Perthes et IH Besser, 1809.

[55] Abdullah Gök, Alec Waterworth, and Philip Shapira. Use of web mining in studying innovation. *Scientometrics*, 102(1):653–671, 2015.

[56] Stephen J. Hanson and Lorien Y. Pratt. Comparing biases for minimal network construction with back-propagation. In *Proceedings of the 1st International Conference on Neural Information Processing Systems*, pages 177–185, 1988.

[57] Donald O. Hebb. The organization of behavior john wiley & sons. *New York*, 1949.

[58] Diana Hicks, Dirk Libaers, and Alan L. Porter. Identification of the technology commercialization strategies of high-tech small firms. *Small Business Research Summary, December*, 2006.

[59] Widrow B. Hoff. Me (1960) adaptive switching circuits. In *IRE WESCON Convention Record*, volume 4, pages 96–104, 1960.

[60] Lis Højgaard. Tracing differentiation in gendered leadership: an analysis of differences in gender composition in top management in business, politics and the civil service. *Gender, Work & Organization*, 9(1):15–38, 2002.

[61] Holger H. Hoos. Automated algorithm configuration and parameter tuning. In *Autonomous search*, pages 37–71. Springer, 2011.

[62] John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[63] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[64] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[65] Andreas Hotho, Andreas Nürnberger, and Gerhard Paaß. A brief survey of text mining. In *Ldv Forum*, volume 20, pages 19–62. Citeseer, 2005.

[66] Jiaoyang Huang and Horng-Tzer Yau. Dynamics of deep neural networks and neural tangent hierarchy. In *International Conference on Machine Learning*, pages 4542–4551. PMLR, 2020.

[67] Michael I. Jordan. Attractor dynamics and parallelism in a connectionist sequential machine. in\fiproceedings of the eighth annual conference of the cognitive science society.\fp hillsdale, 1986.

[68] Shulamit Kahn and Donna Ginther. Women and stem. Working Paper 23525, National Bureau of Economic Research, June 2017.

[69] Helen Kang, Mandy Cheng, and Sidney J. Gray. Corporate governance and board composition: Diversity and independence of australian boards. *Corporate Governance: An International Review*, 15(2):194–207, 2007.

[70] Adnan Khashman. Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, 37(9):6233–6239, 2010.

[71] W. Chan Kim and Renee Mauborgne. Creating new market space. *Harvard business review*, 77(1):83–93, 1999.

[72] W. Chan Kim and Renee Mauborgne. Value innovation: a leap into the blue ocean. *Journal of business strategy*, 2005.

[73] Philipp Koellinger, Maria Minniti, and Christian Schade. Gender differences in entrepreneurial propensity*. *Oxford Bulletin of Economics and Statistics*, 75(2):213–234, 2013.

[74] Michael Kohlgrüber, Karina Maldonado-Mariscal, and Antonius Schröder. Mutual learning in innovation and co-creation processes: Integrating technological and social innovation. *Front. Educ. 6: 498661. doi: 10.3389/feduc*, 2021.

[75] Teuvo Kohonen. Self-organizing feature maps. In *Self-organization and associative memory*, pages 119–157. Springer, 1989.

[76] Teuvo Kohonen. Speech recognition based on topology-preserving neural maps. In *Neural Computing Architectures: The design of brain-like machines*, pages 26–40. 1989.

[77] Oskar Korkman et al. *Customer value formation in practice: a practice-theoretical approach.* Svenska handelshögskolan, 2006.

[78] M.S. Krishnan and Prahalad Coimbatore K. *The new age of innovation.* Tata McGraw-Hill Education, 2008.

[79] Per Kristensson, Jonas Matthing, and Niklas Johansson. Key strategies for the successful involvement of customers in the co-creation of new technology-based services. *International journal of service industry management*, 2008.

[80] Per Kristensson, Jonas Matthing, and Niklas Johansson. Key strategies for the successful involvement of customers in the co-creation of new technology-based services. *International journal of service industry management*, 2008.

[81] Nirmalya Kumar, Lisa Scheer, and Philip Kotler. From market driven to market driving. *European management journal*, 18(2):129–142, 2000.

[82] Adrien-Marie Legendre. New methods for the determination of orbits of comets. *Courcier, Paris*, 1805.

[83] Judith Lorber. The social construction of gender. In *Inequality in the 21st Century*, pages 347–352. Routledge, 2018.

[84] Jyoti D. Mahadeo, Teerooven Soobaroyen, and Vanisha O. Hanuman. Board composition and financial performance: Uncovering the effects of diversity in an emerging economy. *Journal of business ethics*, 105(3):375–388, 2012.

[85] Constantinos Markides. Strategy and management: Constantinos markides discusses strategic innovation. *European Management Journal*, 18(4):357–366, 2000.

[86] Susan Marlow and Maura McAdam. Gender and entrepreneurship: Advancing debate and challenging myths; exploring the mystery of the under-performing female entrepreneur. *International Journal of Entrepreneurial Behavior & Research*, 2013.

[87] Lynn Martin, Bob Jerrard, and Lucy Wright. Identity work in female-led creative businesses. *Gender, Work & Organization*, 27(3):310–326, 2020.

[88] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[89] Toyah Miller and María del Carmen Triana. Demographic diversity in the boardroom: Mediators of the board diversity–firm performance relationship. *Journal of Management studies*, 46(5):755–786, 2009.

[90] Maria Minniti. *Gender issues in entrepreneurship*. Now Publishers Inc, 2009.

[91] Marvin Minsky and Seymour Papert. Perceptrons cambridge. *MIT Press*, 1969.

[92] Michael D. Mumford, Stephen J. Zaccaro, Mary S. Connelly, and Michelle A. Marks. Leadership skills: Conclusions and future directions. *The Leadership Quarterly*, 11(1):155–170, 2000.

[93] Satish Nambisan. Virtual customer environments: It-enabled customer co-innovation and value co-creation. In *Information technology and product development*, pages 109–127. Springer, 2010.

[94] Satish Nambisan and Robert A. Baron. Virtual customer environments: testing a model of voluntary participation in value co-creation activities. *Journal of product innovation management*, 26(4):388–406, 2009.

[95] Satish Nambisan, Kalle Lyytinen, Ann Majchrzak, and Michael Song. Digital innovation management: Reinventing innovation management research in a digital world. *Mis Quarterly*, 41(1), 2017.

[96] Sabina Nielsen and Morten Huse. Women directors' contribution to board decision-making and strategic involvement: The role of equality perception. *European Management Review*, 7(1):16–29, 2010.

[97] Edward Nissan, Inmaculada Carrasco, and María-Soledad Castaño. Women entrepreneurship, innovation, and internationalization. In *Women's entrepreneurship and economics*, pages 125–142. Springer, 2012.

[98] Richard Normann. *Reframing business: When the map changes the landscape*. John Wiley & Sons, 2001.

[99] Alexander Osterwalder and Yves Pigneur. Business model canvas. *Self published. Last*, 2010.

[100] Charles A. O'Reilly III and Michael L. Tushman. Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in organizational behavior*, 28:185–206, 2008.

[101] Adrian F. Payne, Kaj Storbacka, and Pennie Frow. Managing the co-creation of value. *Journal of the academy of marketing science*, 36(1):83–96, 2008.

[102] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[103] Coimbatore K. Prahalad and Venkatram Ramaswamy. The new frontier of experience innovation. *MIT Sloan management review*, 44(4):12, 2003.

[104] Coimbatore K. Prahalad, Venkatram Ramaswamy, et al. Co-opting customer competence. *Harvard business review*, 78(1):79–90, 2000.

[105] Jonathan K. Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.

[106] Kumar R. Ranjan and Stuart Read. Value co-creation: concept and measurement. *Journal of the Academy of Marketing Science*, 44(3):290–315, 2016.

[107] Russell D. Reed and Robert J. Marks. *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. MIT Press, Cambridge, MA, USA, 1998.

[108] Peter Rosa and Alison Dawson. Gender and the commercialization of university science: academic founders of spinout companies. *Entrepreneurship and Regional Development*, 18(4):341–366, 2006.

[109] Frank Rosenblatt. Principles of neurodynamics: Perceptrons and the theory of brain mechanism. *Brain Mechanisms*, pages 555–559, 1962.

[110] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[111] Mohanbir Sawhney, Gianmario Verona, and Emanuela Prandelli. Collaborating to create: The internet as a platform for customer engagement in product innovation. *Journal of interactive marketing*, 19(4):4–17, 2005.

[112] Joseph A. Schumpeter. *49. Capitalism, Socialism and Democracy*. Edinburgh University Press, 2016.

[113] Terrence J. Sejnowski and Geoffrey E. Hinton. Separating figure from ground with a boltzmann machine. In *Vision, brain, and cooperative computation*, pages 703–724. 1990.

[114] Scott Shane and Sankaran Venkataraman. The promise of entrepreneurship as a field of research. *Academy of management review*, 25(1):217–226, 2000.

[115] Charles Spearman. The proof and measurement of association between two things. 1961.

[116] Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[117] Stoyan Tanev, Tony Bailetti, Steve Allen, Hristo Milyakov, Pavel Durchev, and Petko Ruskov. How do value co-creation activities relate to the perception of firms' innovativeness? *Journal of Innovation Economics Management*, (1):131–159, 2011.

[118] Stoyan Tanev, Tony Bailetti, Steve Allen, Hristo Milyakov, Pavel Durchev, and Petko Ruskov. How do value co-creation activities relate to the perception of firms' innovativeness? *Journal of Innovation Economics Management*, (1):131–159, 2011.

[119] Victor A. Thompson. Bureaucracy and innovation. *Administrative science quarterly*, pages 1–20, 1965.

[120] Franz Tödtling, Patrick Lehner, and Alexander Kaufmann. Do different types of innovation rely on specific kinds of knowledge interactions? *Technovation*, 29(1):59–71, 2009.

[121] Andrew H. Van de Ven. Central problems in the management of innovation. *Management science*, 32(5):590–607, 1986.

[122] Guido Van Rossum et al. Computer programming for everybody. *Proposal to the Corporation for National Research initiatives*, 1999.

[123] Stephen L. Vargo and Robert F. Lusch. The four service marketing myths: remnants of a goods-based, manufacturing model. *Journal of service research*, 6(4):324–335, 2004.

[124] Stephen L. Vargo and Robert F. Lusch. Service-dominant logic. *The service-dominant logic of marketing: Dialog, debate, and directions*, 2006.

[125] John Von Neumann. Papers of john von neumann on computers and computer theory. 1986.

[126] Judy Wajcman. Feminist theories of technology. *Cambridge journal of economics*, 34(1):143–152, 2010.

[127] John Watson and Mark McNaughton. Gender differences in risk aversion and expected retirement benefits. *Financial Analysts Journal*, 63(4):52–62, 2007.

[128] John Watson and Rick Newby. Biological sex, stereotypical sex-roles, and sme owner characteristics. *International Journal of Entrepreneurial Behavior & Research*, 2005.

[129] Paul Werbos. Backpropagation: Past and future. In *Proceedings of the Second International Conference on Neural Network*, volume 1, pages 343–353. IEEE, 1988.

[130] Melissa G. Witt and Wendy Wood. Self-regulation of gendered behavior in everyday life. *Sex Roles*, 62(9-10):635–646, 2010.

[131] Stephen J. Zaccaro, Andrea L. Rittman, and Michelle A. Marks. Team leadership. *The leadership quarterly*, 12(4):451–483, 2001.

[132] Shaker A. Zahra and Jeffrey G. Covin. Domestic and international competitive focus, technology strategy and company performance: An empirical analysis. *Technology Analysis & Strategic Management*, 6(1):39–54, 1994.

[133] Xiang Zhang and Rongqiu Chen. Examining the mechanism of the value co-creation with customers. *International Journal of Production Economics*, 116(2):242–250, 2008.