Universitat de Lleida

ICT Tools to analyze social media and be applied in the business field

Adrià Pons Vilalta

Universitat de Lleida

TESI DOCTORAL

ICT Tools to analyze social media and be applied in the business field

Adrià Pons Vilalta

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida
Programa de Doctorat en Dret i Administració d'Empreses

Director/a
Carla Vintró Sanchez
Josep Rius Torrentó

Tutor/a
Josep Rius Torrentó

ICT TOOLS TO ANALYZE SOCIAL MEDIA AND BE APPLIED IN THE BUSINESS FIELD

by

ADRIÀ PONS VILALTA

Submitted to the department of Law and Business Administration
in partial fulfillment of the requirements for the degree of
PhD Thesis in Business Administration and Finance information
at the

UNIVERSITY OF LLEIDA

May 2022

Thesis supervisors:

Dr. Carla Vintró Sanchez
Associate Professor

Dr. Josep Maria Rius Torrentó
Associate Professor

ICT TOOLS TO ANALYZE SOCIAL MEDIA AND BE APPLIED IN THE BUSINESS FIELD.

by

Adrià Pons Vilalta

Submitted to the department of Law and Business Administration in partial fulfillment of the requirements for the degree PhD Thesis in Business Administration and Finance information

Abstract

Background: In today's world the data is considered as an extremely valued asset and its volume is increasing exponentially every day. Social media networks have become one of the most important data source for collecting large amount of data. This extreme growth brings new challenges and opportunities to individuals and companies with which, through big data and artificial intelligence technologies can capitalize this knowledge in different business areas.

Methods: Different software frameworks have been developed in order to manage big text data in the study of CSR and sentiment analysis in the mining and leather industries to classify its mood in twitter. A structured survey questionnaire to collect data from multiple stakeholders was used as a state of the art for the leather industry. Finally, a set of regression models were implemented to develop an intelligent automated method and provide some prior knowledge of the stocks' likely risk levels.

Results: This thesis explores techniques to extract, manage and analyze tweets and its perceptions from the CSR perspective. Results in the twitter sphere are mainly centered on the impact of mining industries on the land and on ethical concerns as well as related with the effects of leather industries against the planet and the economy. Furthermore, the IFM method proved that can properly predict the risk of securities when there are no drastic events.

Conclusions: A collection of ICT tools to gather and analyze data from social media using big data and AI were implemented. The outcome led to the recommendation to improve CSR disclosure based on social media due to the low percentage of tweets written by official accounts in both leather and mining industries. Results also revealed a slight tendency to positive (48%) tweets about mining in front of negative (43%) related to leather. Even though the IFM method could evaluate the risk of equities, a new approach is proposed to reach its full potential.

Resum

Introducció: Actualment, les dades es consideren un actiu molt valorat i el seu volum augmenta de manera exponencial cada dia que passa. Les xarxes socials s'han convertit en una de les fonts de dades més importants per recopilar una gran quantitat d'informació. Aquest creixement extrem aporta nous reptes i oportunitats a persones i empreses amb les quals, a través del big data i les tecnologies d'intel·ligència artificial, podem capitalitzar aquest coneixement en diferents àrees de l'àmbit empresarial.

Mètodes: S'han desenvolupat diferents softwares per tal de gestionar grans quantitats de texts en l'estudi de la RSC i l'anàlisi de sentiments a les indústries de la mineria i la pell per classificar el seu estat d'ànim a twitter. S'ha utilitzat un qüestionari per recollir dades de les múltiples parts interessades com a estat de l'art per a la indústria del cuir. Finalment, s'ha implementat un conjunt de models de regressió per desenvolupar un mètode automatitzat intel·ligent i proporcionar un coneixement previ dels nivells de risc probables de les accions.

Resultats: Aquesta tesi explora tècniques per extreure, gestionar i analitzar els tuits i les seves percepcions des de la perspectiva de la RSC. Els resultats es centren principalment en l'impacte de les indústries mineres al sòl i en qüestions ètiques, així com en els efectes de les indústries del cuir contra el planeta i l'economia. A més, el mètode IFM ha demostrat que pot predir correctament el risc de les accions borsàries quan no hi ha esdeveniments dràstics que les afectin.

Conclusions: S'han implementat una col·lecció d'eines TIC per recopilar i analitzar dades de les xarxes socials mitjançant big data i IA. Els resultats recomanen millorar la divulgació de la RSC basada en les xarxes socials a causa del baix percentatge de tuits escrits per comptes oficials tant a la indústria del cuir com a la mineria. Els resultats també revelen una lleugera tendència de tuits positius (48%) sobre mineria davant els negatius (43%) relacionats amb la pell. Per altra banda, el mètode IFM permet avaluar el risc de les accions, tot i que es proposa un nou enfocament per assolir tot el seu potencial.

Resumen

Introducción: Actualmente, los datos se consideran un activo muy valorado y su volumen aumenta de forma exponencial todos los días. Las redes sociales se han convertido en una de las mayores fuentes de datos para recopilar una gran cantidad de información. Este crecimiento extremo aporta nuevos retos y oportunidades a personas y empresas con las que, a través del big data y las tecnologías de inteligencia artificial, podemos capitalizar este conocimiento en distintas áreas del ámbito empresarial.

Métodos: Se han desarrollado diferentes softwares para gestionar grandes cantidades de textos para estudiar la RSC y el análisis de sentimientos en las industrias de la minería y la piel para clasificar su estado de ánimo en twitter. Se ha utilizado un cuestionario para recabar datos de las múltiples partes interesadas como estado del arte para la industria de la piel. Por último, se ha implementado un conjunto de modelos de regresión para desarrollar un método automatizado inteligente y proporcionar un conocimiento previo de los niveles de riesgo probables de las acciones.

Resultados: Esta tesis explora técnicas para extraer, gestionar y analizar los tuits y sus percepciones desde la perspectiva de la RSC. Los resultados se centran principalmente en el impacto de las industrias mineras en el suelo y en cuestiones éticas, así como en los efectos de la industria de la piel contra el planeta y la economía. Además, el método IFM ha demostrado que puede predecir correctamente el riesgo de las acciones bursátiles cuando no hay eventos drásticos que les afecten.

Conclusiones: Se han implementado una colección de herramientas TIC para recopilar y analizar datos de las redes sociales mediante big data y IA. Los resultados recomiendan mejorar la divulgación de la RSC basada en las redes sociales debido al bajo porcentaje de tuits escritos por cuentas oficiales tanto en la industria del cuero como en la minería. Los resultados también revelan una ligera tendencia de tuits positivos (48%) sobre minería frente a los negativos (43%) relacionados con la piel. Por otra parte, el método IFM permite evaluar el riesgo de las acciones, aunque se propone un nuevo enfoque para conseguir todo su potencial.

Acknowledgements

During this PhD period I have gone through different situations which have made me leave my comfort zone and thus enrich my knowledge and experience in certain areas that I could have never learned anywhere else. Throughout all stages from this PhD I have received a great deal of support and assistance.

I want to thank my supervisors, Dr. Carla Vintró and Dr. Josep Maria Rius, for the much needed guidance and support. Your commitment without expecting anything in exchange has been inspiring for me as a newcomer to the academic and research world.

I would like to thank all the co-authors members of the Department of Computer Science and Business Administration, at the University of Lleida, Oriol Querol, Jordi Vilaplana and Cristobal Fransi for being generous in your effort and collaboration.

I would also like to thank my friends and family for their inconditional support throughout these last years, specially my mother, Emilia and my father-in-law Francesc, who have transmitted me the values of the University of Lleida and its academic community as well as the emphasis on research. To my father Josep, for his constant willingness to support. And in particular, I must acknowledge my wife, Sara, for her encouragement and guidance without exceptions.

Adrià Pons

Lleida, Catalonia

May, 2022

Table of Contents

Chapter 1

Introduction and scope of the research

This thesis presents a collection of techniques to take advantage of massive data gathered from social media by using Big Data and Artificial Intelligence (AI) with the aim of applying this knowledge in different areas from the corporate world. On the one hand, it is intended to analyze stakeholders' perceptions from the corporate social responsibility (CSR) perspective of two specific industrial sectors: mining and leather tanning. On the other hand, it is proposed a sentiment analysis model that anticipates risk of stock market shares taking into consideration unexpected market events.

Over the course of centuries, stakeholders and companies' decision makers have demanded more and more data to support their decision-making process in their daily life. The improvement of database management and data warehousing technologies together with cloud computing and Internet of Things (IoT) has made grown big Data technology and its analysis, to the point where it touches nearly every aspect of our lifestyles, shopping habits, and routine consumer choices. Several industries have embraced the process of research into massive amounts of data to reveal hidden patterns and secret correlations named as big data analytics and thus gain richer and deeper insights to get advantage over the competition. But also opens new opportunities for industries that are seeking for sustainability and ethical management of their activities by gathering data from social media to know and grasp their interest groups, enhance dialogue, meet their demands and expectations as well as be transparent when reporting back their actions to generate real value for their different stakeholders and contribute to the development of CSR strategies.

The arrival of big data has shown an improvement in the identification and analysis of perceptions and sentiments from stakeholders which are relevant to all sectors, but even more for those seen as controversial industries. Given that now, the discussions have mainly shifted to social media where either internal and external stakeholders express most of their comments and criticisms in relation to actions taken by companies and its consequences, CSR communication has become the key to connect the objective key results of companies with society and stakeholders' expectations.

This goal can be achieved by means of different techniques and approaches that are dealt with in this work applied to three specific industries: Mining, Leather and Finance.

Figure 1.1 shows an overall schema of the work involved in this thesis where data has been gathered from different sources, stored and processed using different forms of AI for a further analysis.

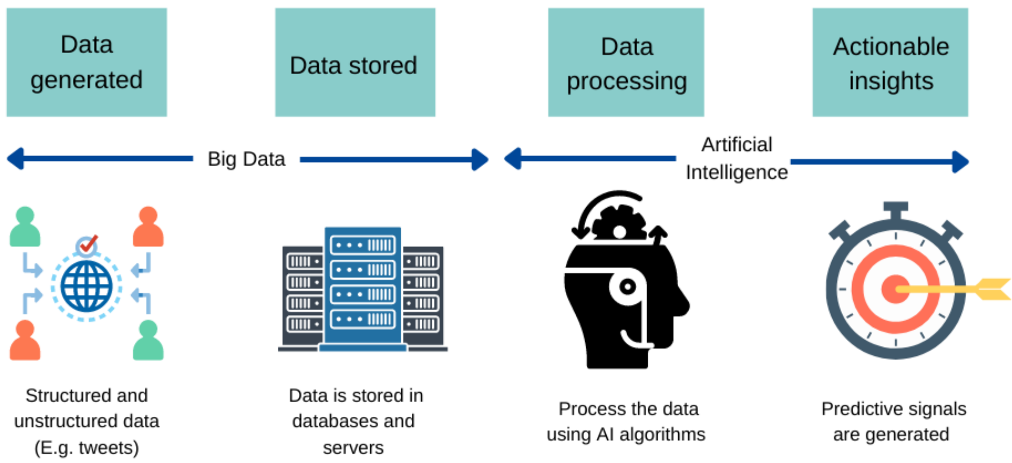


Figure 1.1: Generic big data and AI framework

New insights gleaned from such data-value extraction can meaningfully complement official statistics, surveys, and archival data sources that remain largely static, adding depth and insight from collective experiences and doing so in real time, thereby narrowing both information and time gaps.

1.1 Big data in the Business Field

In today's world ICT and innovation are the pillars of business growth and big data has become an important factor in success of organizations. For this reason data is considered to be the most prominent asset for companies. Enterprises across the globe, from big to small and medium (SMEs), are exploring new ways to capitalize data. Big Data usage is not just for multinational organizations. Currently, SMEs, too can take advantage of the enormous amount of data to take fast and accurate decisions to improve their several business areas. Various Scholars (Beaver et al. 2010) (Wamba et al. 2015) (Davenport 2014) reported that big data is a paradigm shift to improve the business processes, so there is a desperate requirement to seriously think about big data adoption.

The abundant research in academia and industry shows that even retailers can achieve up to 15–20% increase in ROI by putting big data into analytics (Perrey, Spillecke, and Umblijs 2013). (Brown, Chui, and Manyika 2011) found that collecting, storing, and mining big data for insights can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating a substantial economic surplus for consumers. In addition, 'big data' has the capability of transforming the decision making process by allowing enhanced visibility of firm operations and improved performance measurement mechanisms (McAfee et al. 2012).

Over the past few years, big data research has become highly diversified, and has yielded a number of published studies supported by machine-based processes such as text analysis (quantitative linguistics), sentiment analysis (mood recognition), social network analysis, or image analysis, or other processes of a machine-based nature that with some creative tweaking, it might be used to provide new insights to classic business questions like: What offers should we present to a customer? Which customers are most likely to stop being customers soon? How much stock should we hold in the warehouse? How should we define prices of our products? But big data also offers a promising new dimension. The competitive nature of business is heavily depending on the real-time data analysis and capability to predict future outcomes and trend even by a fraction of minutes might just be the differentiating factor that enables to gain competitive advantage. The continuing stream of big data suggests that organizations need to think about new ways of making decisions with this resource instead of just creating reports or presentations that advise senior executives on internal decisions.

1.2 CSR

1.2.1 *Definitions*

Up until now there is no standard definition of the term ”corporate social responsibility” which some also refer to as ”corporate responsibility” or ”Corporate Citizenship”. Indeed, several associated definitions might be found.

For example (Galán 2008) defines CSR as a comprehensive model of business management focused on satisfying needs and expectations of the companies' interest groups and the care and preservation of the environment.

(Martínez and Cuesta González 2003) defines it as the recognition and integration of social and environmental concerns into enterprise operations, giving rise to business practices that satisfy these concerns and that help to build relationships with their stakeholders.

Other authors refer to it as a powerful management tool (Carroll 1999) integrated into the companies' strategy, mission and values compatible with the traditional business vision (guided by profitability management) and social or environmental actions typical from the sustainable development vision.

Green book of the European Union refers to this concept as the method whereby companies integrate social and environmental concerns in their business operations and in their interaction with their stakeholders on a voluntary basis. It is about enterprises deciding to go beyond minimum legal requirements and obligations stemming from collective agreements in order to address societal needs (Commission 2006)

World Business Council for Sustainable Development (WBCSD) provides a definition that might be considered as a global description: ”Is the continuing commitment by business to behave ethically and contribute to economic development while improving the quality of life of the workforce and their families as well as of the local community and society at large.”

Despite CSR has become a modern corporate mantra there is still an absence of a consensus around a definition which might be a major debility of the field, however the International Organization for Standardization (ISO) launched in 2010 after some years of negotiations among several stakeholders across the world a standard that defines CSR as the responsibility of an organization for the impacts of its decisions and activities on society and the environment, through transparent and ethical behaviour that:

- Contributes to sustainable development, including health and the welfare of society
- Takes into account the expectations of stakeholders
- Is in compliance with the applicable law and consistent with international norms of behaviour and
- Is integrated throughout the organization and practiced in its relationships

This standard may help businesses and organizations to translate principles into effective actions and shares best practices relating to social responsibility, globally. It is aimed at all types of organizations regardless of their activity, size or location (ISO 2010).

1.2.2 *Background and historical perspectives*

CSR is far from being a new concept. Some date back its origins to the 19th century due to the existence of exemplary businessmen, namely Owen or Cadbury, although some references might be found at a much earlier stage, such as the one cited by Cicero in the year AD 50 in which he proclaimed *"justice is essential for conducting business"* (Bestratén and Pujol 2003).

Since the 1950s, there has been a “great acceleration” in the human imprint on the natural world. There have been dramatic shifts in key earth systems trends, namely emissions of greenhouse gases (GHG), and socio-economic trends as urban population growth (Steffen et al. 2015). These facts had helped to increase attention on global sustainable development at the beginning of the 70s.

In 1972, the economist Paul A. Samuelson in a text entitled "Economics from the heart" made the following statement in relation with CSR (Redondo 2006):

"It is clear that over the coming years the so-called large private sector will be subject to external limitations that Harvard Business School have never ever imagined. Chairmen will face difficulties when following policies that will cause pollution to the atmosphere. In addition, they will discover that hundreads of traditional business decision-making processes won't be allowed. Society will make companies bear more responsibility and will take action to make it comply[...]. It is therefore interesting to experience new development ways of the most prosperous nations. Higher social responsibilities to companies will be requested."

Between 70s and 80s the link between reputation, management and community outreach was narrower (Romera 2006) and, thus, social responsibility gained

Period & Focus Area	Summary of Dimensions
1950's – 1960's	
• Religious & Humane philosophies	
• Community development	
• Unregulated philanthropy	Philanthropy
• Poverty alleviation	
• Obligation to the society	
1970's – 1980's	
• Extension of CSR commitments	
• CSR as symbol of Corporate citizenship	
• Stakeholder relationship management	
• Corporate reputation	Regulated CSR
• Socio-economic priorities	
• Bridging governance gap	
• Stakeholders rights	
• Legal & Ethical responsibilities	
1990's – 21st Century	
• Competitive strategy	
• Environmental protection	
• Sustainability	Instrumental/Strategic CSR
• Internationalisation of CSR standards	
• Transparency & accountability	

Table 1.1: Evolution of CSR perspective

importance in different business practices.

Nevertheless, it has been in the new millennium when CSR has increased its popularity and has become a trend topic (Antolín and Gago 2004) formalized as a key business driver with the adoption of the Paris Agreement and the Sustainable Development Goals (SDGs) adopted in 2015, which reflected a new social contract in which corporations are expected to play a relevant role in the global efforts to achieve the SDGs, however they do not represent any commitments for the private sector, the countries that adopt them will have to create specific policies and regulations that will translate into pressure for companies to implement new business practices or to improve their current ones. This context presents an opportunity for CSR to continue growing in terms of conceptualization and implementation, mainly because businesses can adopt it as a strategic framework with the objective of creating shared value.

Presently, the pandemic has led to a sharp increase in governments' and market participants' attention to CSR considerations. For instance, social and environmental issues are at the core of the recovery plan in many countries. The European Parliament recommitted to the European Green Deal, a set of policy initiatives introduced in December 2019 that aim to make Europe climate-neutral by 2050. It is also trying to build post–COVID-19 economic stimulus packages around the goals of this Green Deal (Bae, El Ghoul, and Zhaoran (Jason) Gong 2021).

The table below summarises the historical scope evolution. (Rahman et al. 2011).

1.2.3 *Stakeholder engagement*

One of the current aims of CSR is to connect the objective key results of the company with society and stakeholders' expectations. These expectations can be very diverse due to according to (ISO 2010) stakeholders are people or groups who are affected by the actions of an organization. This includes workers, clients, purchasers, consumers, owners, investors, government officials, community residents, and suppliers. For example, a hospital's stakeholders could include doctors, nurses, patients, patients' families, the owners of the hospital (could be a branch of government, or private investors), the community where the hospital is located, the suppliers of medical items, etc. Because of the financial, physical and environmental impacts that businesses can have on individuals and communities, it is vital that businesses essentially identify and communicate with those people who have some legitimate stake in them (Noland and Phillips 2010). Thus and in compliance with (O'Riordan and Fairbrass 2014), stakeholder engagement refers to a variety of activities[1] that enable dialogue between an organization and its stakeholders.

By engaging in CSR activities, companies can not only generate positive stakeholder attitudes and better supportive behaviours (e.g. purchase, seeking employment, investing in the company) but also, build a strong and credible corporate image, strengthen stakeholder company relationships, and enhance stakeholders' advocacy behaviours in the long-term. An effective communication of CSR actions to stakeholders not only increases their level of credibility and involvement with the company but also turns them into the company's advocates on social media.

It is therefore clear the impetus behind the use of the term 'engagement' together with stakeholder and CSR is the need to emphasize that, for companies merely to interact in an unidirectional way with stakeholders is no longer enough, if, in fact, it ever was. However, it is possible to trade with another actor without ever engaging him or her as a fellow person; that is, transacting without inquiring as to his or her wants, needs, well-being or capabilities. With this in mind, 'engagement', is used to recommend a type of interaction that involves, at minimum, recognition and respect of common humanity and the ways in which the actions of each may affect the other.

Consequently, knowing stakeholders as individuals with names and faces might be among the best and most concise ways of characterizing the essence of stakeholder engagement, but to reach this level of depth can be complicated. Nonetheless,

[1] Referring to, for example, communication, collaboration, consultation, dialogue, and joint decision-making.

is important to don't abandon this objective because for a company to determine its strategy without having first engaged its stakeholders and know more about their sentiments towards the firm would be, literally, to disengage its mission and vision from its identity.

1.3 Social Media

It is vital to analyze the real composition of the audiences and their expectations from the firm's activities and identify a useful tool for increasing the dialogue (Iazzi et al. 2020) (Du and Vieira 2012). According to the dialogic principle, the dialogue is defined as any negotiated exchange of ideas and opinions characterized by an honest, open, and ethically-based give and take (Bortree and Seltzer 2009). In this sense, communication via social media could be a enabler to engage with stakeholders (Benthaus, Risius, and Beck 2016).

Social media can be understood as "Internet-based applications that carry consumer-generated content which encompasses media impressions created by consumers, typically informed by relevant experience, and archived or shared online for easy access by other impressionable consumers (Xiang and Gretzel 2010). Therefore, people can use this media to create content, or freely express their opinions about companies' brands, or behaviours. It is also being used by corporations to cultivate relationships with key stakeholders (Kelleher 2006) and engage in conversations since it allows for rapid dissemination and exchange of information within minimum time and cost on a massive scale.

Media activities generated by stakeholders or communities that are neither paid for nor initiated by brand owners are claimed to have a potentially game-changing impact on communication and building corporate reputation (Ali, Jiménez-Zarco, and Bicho 2015). The communication through social media is more transparent and helpful in shaping stakeholders behaviour including that of consumers. The emerging developments in communication technologies and the IT revolution in recent years have forced corporations to use new communication strategies to convey their important messages to engage stakeholders. Traditional communication channels and strategies are relatively less effective in today's digital age. They need to engage in consumer's conversations. The classic notion of a 'target customer' in the companies-communication paradigm must be enriched to take account of the fact that the consumer now has a voice and wants to be heard. And social media gives the consumers the opportunity to voice their opinions and be heard.

Popular and well-known social media platforms such as Facebook, Instagram, YouTube, and Twitter also make it easy for users to exert influence on products or companies. User-generated content, in the form of text, pictures, and videos, creates a data-rich environment in which companies can gather feedback and create competitive intelligence. However, their ubiquity and complexity (e.g., vast amounts of data, rapid information diffusion, interactivity, and reach) make it challenging for firms to take advantage of it.

1.3.1 *Twitter*

Social media cover a wide range of Social Network Sites, that is, sites driven by user-participation and user-generated content. From those, Twitter which was launched on July 13, 2006, has become one of the most popular micro-blogging site on the Internet. Around 316 million daily active users over the globe post 500 million tweets per day (Kühl et al. 2019). Tweets are understood as periodic status updates, that consist of short messages of maximum size 280 characters. These updates typically consist of personal information about the users, news or links to content such as images, video and articles. Posts made by a user are displayed on the user's profile page, as well as shown to his/her followers. Users may select other users to "follow", repost ("retweet") updates from that users, pose comments (@reply), and mark "likes".

In 2012 Twitter launched a streaming API[2] which allows to receive tweets and replies in real time from Twitter. For big data research, application programming represents a central tool of data driven knowledge production. The term "application programming interface" is abbreviated here as API. It is an irony of history that the usability of the commercially motivated Twitter API has given a boost to big social data research in the field of media and communication sciences. Through the different access points of the Twitter API, returns all public tweets that correspond to a set of filter parameters passed containing data encoded using JavaScript Object Notation (JSON). In addition to the text content itself, a tweet can have over 150 attributes associated with it, such as; author data, a timestamp of when it as posted, geo metadata shared by the user, hashtags, mentions, etc... Figure 1.2 shows an example of a tweet:

And this is how would be received from the Twitter's API containing just some attributes as an example:

[2]https://developer.twitter.com/en

Figure 1.2: Example of a tweet


```
{
 "created_at": "Sun Apr 10 13:17:24 +0000 2022",
 "id": 1050118621198921728,
 "text": "This is of course a small sample out of the huge dictionary
 composing each tweet. ",
 "user": {
    "id": 2244994945,
    "name": "Adria Pons",
    "location": "Internet",
  },
 "entities": {
    "hashtags": [
    ],
 "retweeted": false,
 "retweet_count": 0,
}
```


This opens a new window to build models to aggregate opinions of the collective population and gain useful insights into their behavior. Most of the user-generated messages on twitter are textual information and identifying their sentiments has become an important issue (Liang and Dai 2013).

1.3.2 *Sentiment Anaysis*

As mentioned above, Twitter generates massive amounts of data that expresses thoughts and opinions of its users on issues. In this context, private individuals become the sources of information through online sharing of opinions and thoughts.

These thoughts and opinions can be extracted to show patterns on sentiments of its users. Sentiment analysis, also referred as opinion mining, studies the "opinions, attitudes and emotions" of people. Sentiment analysis is a text classification problem, and identifying and extracting the right sentiment from huge data on social media sites are a challenge.

Huge unstructured data is available in many forms like tweets, reviews or news articles etc. which can be classified as positive, neutral or negative polarity according to the sentiment that is expressed by them. There are many challenges in Sentiment Analysis. In the first place, an opinion word which is considered to be positive in one state may be considered negative in a different situation. In addition, people may not always express opinions in the similar manner. Eg: "the offer was a great one" differs completely from "the offer was not so great". Opinion of people may be contradictory in their statements. It is even difficult for a machine to analyze. Most of the time people find it difficult to understand what others mean within a short sentence of text because it lacks context.

Here is where Artificial Intelligence (AI) come into play. As can be seen in figure 1.3, the usual construction process of a text classifier system based on machine learning consists of several sequential phases. First, it is necessary to prepare the data to train the algorithms. This question is crucial when we talk about messages extracted from social networks since it is very common to find messages with spelling mistakes, repetitions of characters, mixed uppercase and lowercase letters. Once the message normalization process is completed, the data is transformed into numerical features that can be processed while preserving the information of the original message. This provides better results than applying machine learning directly to the raw data. And finally the classifier rates the tweets according to their positivity.

1.4 Applied cases: Mining, Leather and Finance

In this thesis, three applied cases were researched in different industries. In the following subsections are described in detail.

1.4.1 *Mining*

Mining industry plays a fundamental role in modern society and in industrial processes, since it provides raw materials and sources of energy. However, obtaining these mineral resources causes an impact and therefore a footprint on the environment. For this reason companies have to deal with the compatibility between the

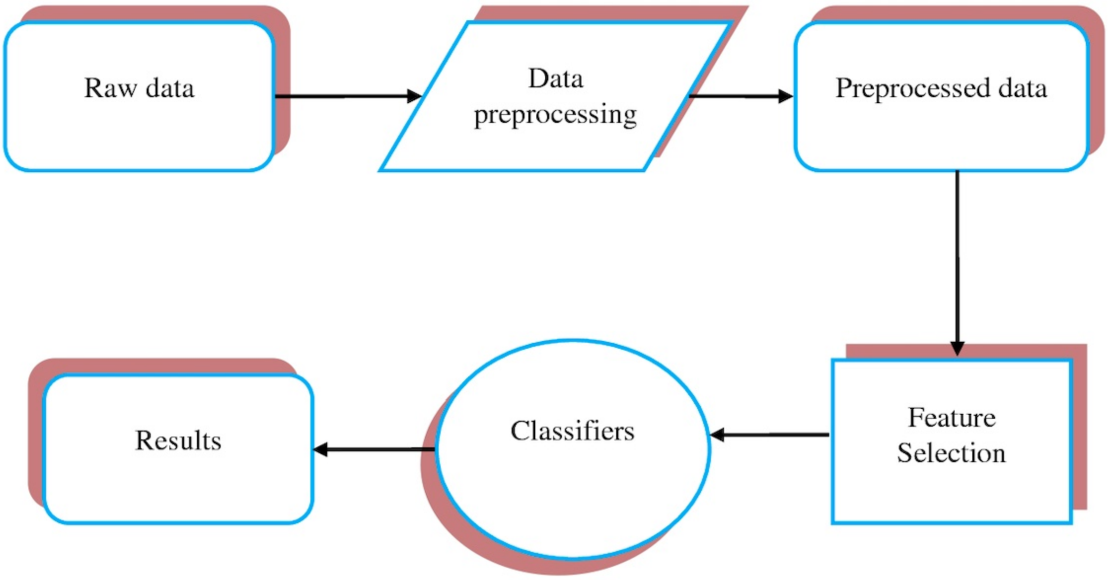


Figure 1.3: Process of Machine learning tasks for sentiment classification

productive activity and Corporate Social Responsibility. The application of CSR practices is not done in a systematic manner, and although social and environmental objectives in company strategies is beginning to gain popularity, most companies do not have a consolidated system of indicators that allow to measure the results in terms of ethical and sustainable management. Hence, it is required to measure the effect of the actions conducted by the companies together with interpretations and feelings generated towards each interest group having greater understanding into the true purpose of the organization, in other words justify the value and impact of the activity on society, introduce more sustainable practices to tackle current challenges and stakeholders' interests, identify essential products and services that define the core business and generate real social value. This would yield useful information about which are the strategic priorities that companies should face from the CSR point of view.

Under this context, this thesis aims to examine CSR communication in the mining sector on Twitter and identify the main topics of CSR and the main participants in the creation of content. To achieve this, 2.000.000 Tweets were gathered using an API to manage big text data in the study of CSR and sentiment computational content analysis to classify the tweets as positive, negative or neutral. To this aim, we have developed a tool in Phyton, through the API of Twitter, the Phyton language, multiple scripts and MongoDB.

1.4.2 *Leather*

A similar approach has been adopted in Leather tanning industry. It is under those sectors historically considered by the society as a threat to the natural surroundings, with environmental effects. Perceived as dirty job, nasty and without any social reputation from a hard and difficult sector which starts from an awkward but inevitable fact; the death of the animal. These characteristics influenced its industry personality making them reticent and unused to communication. This lack of communication from the industry towards its stakeholders has led to consider it one of the most polluting industries in the world due to the use of significant quantities of chemicals, energy and water during leather production process (Kanagaraj et al. 2015) without appreciating the huge industrial transformation and its fundamental role to upcycle an unavoidable waste from the food industry.

Even today, despite having highly valued features and properties, leather is just a product; a bag, a sofa, a jacket, but seems to have no value in itself. The skin goes straight from the slaughterhouse to the Hermès shop window. And everything in between is the absolute darkness. Fashion industry is the main customer (and probably "The savior"), due to this reason the sector has left in hands of the big fashion brands all the image work of its leather.

In order to set lights of consciousness, CSR communication from leather industry has increased over the years and now many companies communicate about economic, environmental and social issues in some form, and most of them engage in sustainability reporting.

Therefore, there is the need to develop tools to measure the impact of the actions undertaken and consequently be able to redefine priorities and responsibilities within the organization from a triple perspective: economic, social and environmental.

We have developed an interactive automatic system which predicts the sentiment of the review/tweets of the people posted in twitter using an automatic micro-service completely independent hosted on a server designed to be integrated inside a back-end ecosystem. As a state of the art we have also analyzed the relation between the application of management systems and the introduction of CSR practices in the leather industry together with their current CSR procedures and measurement systems.

1.4.3 *Finance*

Before crossing a street, we evaluate the risk to have an accident, making sure about the status of the traffic light, but also looking from one side to the other to avoid the consequences of distracted drivers with their mobile phones. When buying a house we verify its year of construction, its outstanding charges or its neighbors to avoid the risks derived from an old-fashioned water system, a pending mortgage or wrong neighbors. It would be ideal to measure risks before leaving home or before buying a new house.

Risk prediction measures can be applied in several contexts. Insurance companies are interested in how high claim liabilities might be. Banks care about profit and loss, specifically how large losses might be. Or for instance, investment professionals seek to describe the probability distribution of an individual asset or portfolio of assets, and worry about return outcomes that may be especially low.

Furthermore, traditional financial theories consider that investors operate rationally in making financial decisions (Kiymaz, Öztürkkal, and Akkemik 2016) and evaluate possible alternatives on the basis of utility and associated risk. However, different studies reported that investors usually do not make rational or logical decisions (Nicolosi, Peng, and Zhu 2009) (Barber and Odean 2001). They have limits to their self-control, and are influenced by their own biases, meaning that they tend to make poor decisions about their investments.

Therefore, risk prediction measures might be needed for both internal purposes, as within a corporate risk management department that informs executives of exposure pockets, or for external purposes, as a pragmatic tool for investors to reduce biases and give some prior knowledge of the likely risk scenarios.

It is because of these reasons we have studied with the prediction of VaR (Value at Risk) the percentage of loss that a financial product would have in the future, to assess the risks and determine the potential loss of equities, thus reducing reasoning influenced by feelings for bank and financial firms seeking to deploy AI and advanced automation in their IT solutions.

We have based our model on the IFM method, which was introduced by (Xu 1996) and commonly used in the financial industry. It is one type of regression model that takes no more input than the time series of the equities, that is, sequence of prices in the market of a given equity, in chronological order. We measured the effectiveness of this model by using two consecutive time intervals of 3 years daily stock closing prices of 4 companies; BMW, Tesla, Samsung and Facebook considering two different scenarios (in normal market conditions and in specific market events that changed the course of securities). The obtained results justify further development of the method to take into account not only the past of the financial product, but also the present, which it is assumed to be the most determining for drastic changes in the market.

Chapter 2

Methodology

2.1 Paper 1: Impact of Corporate Social Responsibility in mining industries

A method has been developed to stream tweets from Twitter related with CSR and Mining industry. As mentioned before the Twitter Streaming API allows to receive tweets and notifications in real time from Twitter. However, it requires a high performance, persistent, always in the connection between the server and Twitter. To tackle this issue, a set of techniques have been applied for tweet streaming 24 h a day, every day of the week, to make sure we do not lose any tweet. The technologies used to this aim have been the programming language Python and MongoDB. Also, Naive Bayes sentiment classifier has been applied to these tweets and the results are depicted in different graphs. Scripts are also developed to maintain the tweets collected in the database.

2.2 Paper 2: CSR in leather industries

The aim is provide an interactive automatic system which predicts the sentiment of the review/tweets of the people posted in twitter linked with the leather industry using an automatic micro-service. The work can be broadly separated into two processes, the first receives data and stores it, the second calculates and writes in graphics the results. Only one social network source of information is processed, notwithstanding the service is thought to allow the integration in a future of more than one incoming source of information. The technologies used for that purpose have been the programming language Python, MongoDB, Mongo-Express and Alpine together with an API powered by Azure (Microsoft) called Text Analytics, which has been used to analyze a text and return the overall sentiment. Some scripts have been developed to manage the data properly.

2.3 Paper 3: Corporate Social Responsibility in the leather tanning sector

A structured survey questionnaire is used to collect data from multiple stakeholders. The research was started with a pilot-test questionnaire that was conducted with in a small number of Catalan leather companies guided by the leather cluster of Barcelona in order to prove viability and to detect difficulties in the interpretation of questions. The final version of the questionnaire comprised 25 items concerning the diffusion of CSR practices, CSR procedures, CSR measurement systems and CSR management systems, and the perception of companies about the benefits experimented in their business results. Some of the items were open-ended questions, others were multiple choice questions and some used a five point Likert scale. All questionnaires were sent by electronic mail (web-based survey tool) and were addressed to the management systems manager or to the business manager.

2.4 Paper 4: An Application of the IFM Method for the Risk Assessment of Financial Instruments

We gathered 3 years of daily stock closing prices. To this data we have applied GARCH fitting to model the time series because of the changes in the variance or volatility over the time and copula fitting for the inter-dependencies. Once we obtained the fitted parameters we performed different monte carlo simulation stages due to its easy implementation and because helped to understand the impact of risk and uncertainty. To calculate the VaR, 1000 paths are used.

Figure 2.1 summarize the implementation of this method.

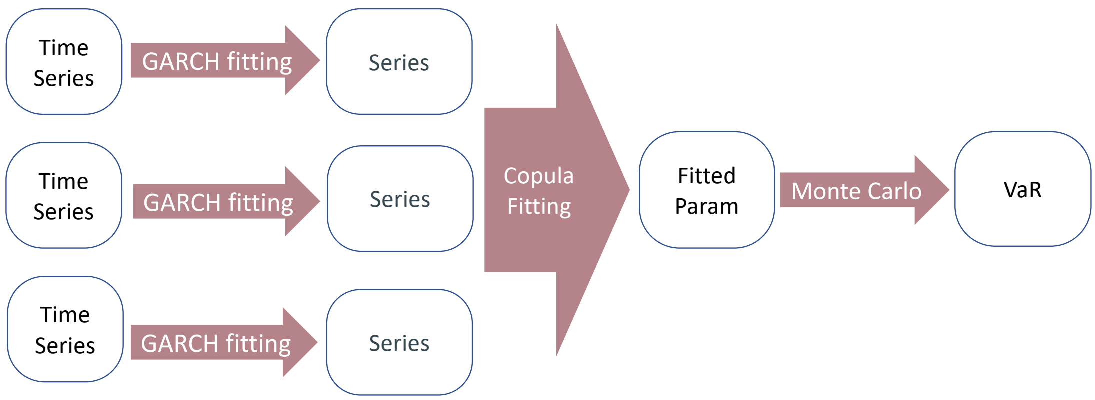


Figure 2.1: Simulation method

Chapter 3

Papers

Resources Policy 72 (2021) 102117

Contents lists available at ScienceDirect

Resources Policy

journal homepage: http://www.elsevier.com/locate/resourpol

Impact of Corporate Social Responsibility in mining industries

Adrià Pons [b], Carla Vintrò [b], Josep Rius [b,*], Jordi Vilaplana [a]

[a] Dept. of Computer Science, University of Lleida, Av. Jaume II 69, 25001, LLeida, Spain
[b] Dept. of Business Administration, University of Lleida, Av. Jaume II 71, 25001, LLeida, Spain

ARTICLE INFO

Keywords:
CSR
Mining
Twitter
Big data
Sentiment analysis

ABSTRACT

The mining sector plays a fundamental role in the global economy since it provides vital raw materials and energy for a large number of industries but mining activities are commonly criticized due to the effects on workers' health and local communities, and are seen as a threat to society in general. For these reasons, companies have to deal with the compatibility between the productive activity and the Corporate Social Responsibility (CSR). Under this context, this paper aims to examine CSR communication in the mining sector on Twitter and identify the main topics of CSR and the main participants in the creation of content. A software framework has been developed in order to manage big text data in the study of CSR and sentiment computational content analysis to classify the tweets as positive, negative or neutral. Twitter was chosen since it is one of the three most commonly used social network sites by mining companies for information sharing and stakeholder engagement. The data collection period was from February 2019 to December 2019 and it resulted in 2,000,000 Twitter posts. The results show the CSR debate is increasingly growing in developing countries and in countries with a bad reputation of environmental and health mining conditions. The debate is mainly centered on the impact of mining industries on the land, and many stakeholders advocate for more environmental responsibility since they consider that mining activities are still controversial.

Also, results reveal that the mining sector should improve the CSR disclosure and adopt a stakeholder engagement strategy grounded in the corporate stakeholder relationship perspective, the two-way symmetrical communication, and the dialogic theory of public relations. Social network sites such as Twitter can lead to positive outcomes for companies since they are not only a way to advertise what the company does with regards to CSR, but also to receive input from other for their CSR activities.

1. Introduction

Over the past several years, stakeholder relationship management has played an important role in strategic leadership and within the scope of business models. Companies that are seeking for sustainability and ethical management of their activities need to know and grasp their interest groups, enhance dialogue, meet their demands and expectations as well as be transparent when reporting back their actions to generate real value for their different stakeholders. Hence, it is necessary to measure the impact of the actions undertaken by the com-panies together with interpretations and feelings being generated towards each interest group, and thus be able to redefine priorities and responsibilities within the organization from a triple perspective: economic, social and environmental. Having greater insight into the real purpose of the organization, in other words justify the value and impact of the activity on society, introduce more sustainable practices to tackle current challenges and stakeholders' interests, identify essential products and services that define the core business and create true social value. This would yield useful information about which are the strategic priorities that companies should face from the Corporate Social Responsibility (hereinafter CSR) point of view. In this regard, environmental impact of economic activities as well as employee security and health conditions are two of the main worldwide concerns and have become an issue of great importance due to are part of the United Nation 2030 agenda for Sustainable Development Goals. These concerns have particular relevance in some sectors depending on the nature of its activities, such as the mining industry.

This sector plays a fundamental role in the global economy since it provides vital raw materials and energy for a large number of industries. Its activities have important economic, environmental, labor and social repercussions on local and global scales Escanciano et al. (2010). As stated by Mendes and Rodrigues, (2018) they have specific

* Corresponding author.
E-mail address: josep.riustorrento@udl.cat (J. Rius).

https://doi.org/10.1016/j.resourpol.2021.102117
Received 21 April 2020; Received in revised form 6 April 2021; Accepted 14 April 2021
Available online 29 April 2021

characteristics because of their transitory nature, and are still commonly considered as a threat to the natural surroundings, with environmental effects on the air, water and soils. Moreover, mining activities are commonly criticized due to the effects on workers' health Sanmiquel et al. (2015) and local communities Raufflet et al. (2014), and are seen as a threat to society in general Mendes and Rodrigues (2018). Further, companies have to deal with the compatibility between the productive activity and the environmental Claver Cortés et al. (2004) and social protection Wheeler et al. (2002), which are pillars of sustainable development.

Research on CSR facilitates a strategy to face all these challenges and helps companies to undertake different actions in order to accept the responsibilities resulting from the impact of its activities on society and environment Vintro et al. (2012) under four dimensions; economic, legal, ethical and philanthropic Carroll (1991). CSR aims to connect the objective key results of the company with society and stakeholders' expectations.

The concept of CSR has gained importance within the academic community and has been an increasing studied subject in recent years, with particular significance in mining activities. As noted by Warhurst Warhurst (2001a), the main environmental disasters and human rights incidents that have increased public concern about CSR over the last 40 years have mainly taken place in the mining and petroleum industries Vintro et al. (2012). The discussion about CSR in the mining industry has emphasized its commitment to social responsibility as an emerging topic Mendes and Rodrigues (2018), and literature seems to indicate that mining companies have increased their environmental and social consciousness Zhu et al. (2010). In this sense, the first decade of the 21st century in particular has seen a renewed debate about mining and its sustainability.

M. Mudd (2010) owing to public concern about the current degradation of the environment Hilson and Murck (2000) and about social actions in the communities affected Hamann (2004). In the literature, there are several studies regarding the extensive interest on the topic of CSR which explore strategies of mining companies on sustainable issues; Sinding (2009); Mclaren et al. (1999); Warhurst and Noronha (2009); Hilson and Murck (2000); Warhurst (2001a); Hilson and Nayee (2002); Newbold (2006); Suppen et al. (2006); Van Zyl et al. (2007); Rodrigues da Silva Enríquez and Drummond (2007); M. Mudd (2010); Fonseca (2010); Dutta et al. (2012); Hassan and Ibrahim (2012); Vintro et al. (2014).

Thus, CSR is increasingly studied by different researchers and is a core strategy of many companies, but as noted by Chae and Park (2018) there are many needs and opportunities for future CSR research, one of those is the application of content analysis to extract themes, categories and sentiments from text corpora. Most of the published studies about content analysis in CSR (see for example Dahlsrud (2008) and Lee and Carroll (2011) have examined small samples of text data from traditional media including journal articles, CSR reports and newspapers, but very few (see for example (Chae and Park, 2018) have examined big amounts of data obtained from social media even though the analysis of big data helps researchers identify some latent knowledge or information to be incorporated in future decision-making Crane and Self (2014). Moreover, these previous studies have mainly used human coding or simple software-based methods in their content analysis (Chae and Park, 2018). When considering the research on CSR in the mining sector, the use of content analysis is even scarcer.

The arrival of big data has shown an improvement in the identification and analysis of perceptions and sentiments from stakeholders which are relevant for all sectors, but even more for those seen as potentially dangerous such as mining. Now, the discussions have mainly shifted to social media where either internal and external stakeholders express most of their comments and criticisms in relation to actions taken by companies and its consequences. There is large interest in studying sentiment analysis from these Stakeholder's comments and posts on social media using big data techniques in order to help companies redefine strategic priorities together with actions to be taken under a social responsibility and environmental framework, which goes beyond the question of economic results.

Under this context, this paper aims to examine CSR communication in the mining sector on Twitter and identify the main topics of CSR and the main participants in the creation of content. In other words, the results obtained will contribute to the development of CSR strategies for the mining industry and similarly other studies may be conducted in different sectors to redefine the CSR role and priorities. The study is sustained in how organizational research focused on social media may translate stakeholder engagement into organizational goals and create the basis of effective strategy development.

The research conducted uses an Application Programming Interface (API) to manage big text data in the study of CSR and sentiment computational content analysis to classify the tweets as positive, negative or neutral. To this aim, the authors developed a tool in Phyton, through the Application Programming Interface (API) of Twitter, the Phyton language, multiple scripts and MongoDB. Twitter was chosen since it is one of the three most commonly used social network sites by mining companies for information sharing and stakeholder engagement. The data collection period was from February 2019 to December 2019 and it resulted in 2,000,000 Twitter posts.

In the next section, the paper explores the current discussions on CSR communication and Social Media. The project architecture of the framework, describing its functionality and the algorithms implemented, is then introduced. Next, the results obtained are presented. Discussions and conclusions follow.

2. Theoretical framework: CSR communication and social media

CSR emphasizes the role of corporate communication in company-stakeholders engagement, according to the principles of transparency and open dialogue with diverse stakeholders to foster ethical and sustainable behavior Lim and Green-wood (2017). In recent years, CSR communication has become a keystone when constructing and sustaining the reputation of a company in the eyes of its stakeholders Türkel and Akan (2015), and it has drawn considerable attention form public relations and CSR researchers Lim and Greenwood (2017). Mining activities might be viewed as more socially and environmentally responsible if CSR practices were adopted. Vintro et al. (2012) concluded that mining companies understand the negative impact of their activities on the environment and they want to collaborate with both internal and external stakeholders.

Until recently, companies mainly communicated through media relations, CSR reports, or websites among other means. Currently, companies also use a wide range of electronic web 2.0 based applications that include social networks, blogs, and photo and video sharing platforms. That is, CSR communication has evolved from one-way communication to two-way communication, and social media has opened new opportunities for CSR related information dissemination and stakeholder relationship management Lattemann and Stieglitz (2007).

Social media can be understood as “Internet-based applications that carry consumer-generated content which encompasses media impressions created by consumers, typically informed by relevant experience, and archived or shared online for easy access by other impressionable consumers” Xiang and Gretzel (2010). Therefore, social media refers to online resources that people use to share content. It is a public channel for firms to engage stakeholders since it allows for rapid dissemination and exchange of information, and by this way enables enterprises to cultivate relationships with key stakeholders Kelleher (2007) and engage in conversations. Stakeholders are no longer simple receivers of information and can now engage in the creation and evaluation of content Dellarocas (2003).

Social media cover a wide range of Social Network Sites, that is, sites driven by user-participation and user-generated content Tredinnick (2006). From those, Facebook, Instagram and Twitter are three of the most commonly used by companies. Specifically, Twitter was launched in October 2006, and it has become the largest micro-blogging site on the Internet. Around 316 million daily active users around the globe post 500 million tweets per day Kühl et al. (2019). It allows users to broadcast real-time messages of 280 characters or less. Users may select other users to “follow”, repost (“retweet”) updates from that users, pose comments (@reply), and mark “likes”. It also functions as a social networking site in that users can connect and share information and is used in official public relations, advertising, and marketing campaigns Stelzner (2012). According to the literature, sentiment classification provides organizations with a tool to transform data into ‘actionable knowledge’ that decision maker can use in pursuit of improved organizational performance. Especially in case of machine learning based, the classification performance can directly depend on the quality of features obtained from a training dataset Cortes et al. (1995) Mitchell (1999) Kira and Rendell (1992). Most of existing sentiment classification studies provide simple performance comparison in terms of the “accuracy” of algorithms without considering the role of “data properties and setting” that may cause differences in the performance, however, recent and advanced theory proposed by Choi and Lee (2017) offers a new approach investigating the impact of linguistic properties of data such as word-count and size of training dataset on the performance of diverse algorithms which leads to the conclusion that the quality and properties of training dataset are of considerable importance for the performance of machine learning and classification accuracy, thus affecting the final results credibility.

Organizational research specific to Twitter has gained importance within the academic community during last years. Different authors have discussed what best practices of communication on Twitter may be boyd et al. (2010), Rybalko and Seltzer (2010), Marwick and Boyd (2011). For instance, Waters and Jamal Waters and Jamal (2011) have examined how non-profit organizations from the Philanthropy 200 communicate on Twitter. Xiang and Gretzel (2010) have analyzed how Twitter contributes to the development of the theory and practice of public relations. Rybalko and Seltzer (2010) have examined how Fortune 500 companies engage stakeholders using Twitter. And similarly, Barnes and E.M (2010) has studied how Fortune 500 companies use Twitter and found that only 35% of those use it, and only 24% of that are actively involved.

Different studies have analyzed organizational engagement through Twitter Andriof and Waddock (2002) Palazzo and Scherer (2006) Smith (2010) Hwang (2013). Kollat and Araujo (2018) state that the need for engaging stakeholders when communicating about CSR activities makes Social Network Sites like Twitter more important, given their ability to foster dialogue and content diffusion. They argue that communicating CSR on Twitter can lead to positive outcomes for companies since it is not only a way to advertise what the company does with regards to CSR, but also to receive input from other for their CSR activities. And Grunig (2009) point that social media make corporation communications more strategic, more interactive and more socially responsible. Similar work has also been pursued by Colleoni (2013) that investigates which corporate communication strategy adopted in online social media is more effective to create convergence between corporations CSR agenda and stakeholders’ social expectations, and thereby, to increase corporate legitimacy applying advanced data-mining techniques, whereas Zhang et al. (2016) explore the online debate in a CSR crisis using Twitter data relating to the recent case of the falsified Volkswagen diesel emissions that became public in 2015 focusing on capturing the issue as it evolved over time, the actors and sentiments expressed, and the responses of the organization. Findings demonstrate that CSR challenges can result in a crisis of a long duration marked by strongly expressed sentiments and a wide diversity in the views of different stakeholder groups.

Twitter can provide a favorable environment for CSR communication, and it seems that organizations with higher CSR ratings tend to build larger communities of followers on Twitter when compared to organizations with lower CSR ratings Lee et al. (2013) Kollat and Araujo (2018). But, while social media is important for CSR communication, the use of social media data is rare in the literature Chae and Park (2018) and more specifically when considering CSR in the mining sector.

Thus, this appears to serve as a reasonable framework for studying the application of sentiment analysis under CSR context where content analysis focused on this industry is scarce. These sentiments are key to go over stakeholders from different perspectives and obtain results in order to adapt CSR priorities as well as its role within mining companies, assuming that mining activities are seen as a threat to society in general.

3. Methodology

3.1. Twitter API

The Twitter Streaming API[1] allows to receive tweets and notifications in real time from Twitter. However, it requires a high performance, persistent, always in the connection between the server and Twitter.

The work of the streaming implementation is to record the incoming events as quickly as possible and process them in the background using the REST API as necessary to collect more in-depth data. The use of the REST API is subject to several speed limits by Twitter. It is important to be a responsible user of the Twitter API by planning limits to the activity within the application and monitoring the API rate limit responses. The streaming API has no speed limits since the data is sent to the server as they appear.

3.2. Streaming tweets

Fig. 1 shows the structure used for the streaming of tweets. The technologies used to this aim were the programming language Python and MongoDB.

MongoDB is a document-oriented database with dynamic schema. This allows it to offer high performance and facilitates the development of applications. In turn, it prevents from having Joins and Transactions, something very common in relational databases. It should be noted that MongoDB stores documents in JSON format and its structures.

Also, Naive Bayes sentiment classifier was applied to these tweets and the results were depicted in different graphs. Scripts were also developed to maintain the tweets collected in the database.

Tweets were streamed in real time with the API of Twitter and a string was created to filter and select the tweets according to the

[1] https://developer.twitter.com/en.html.

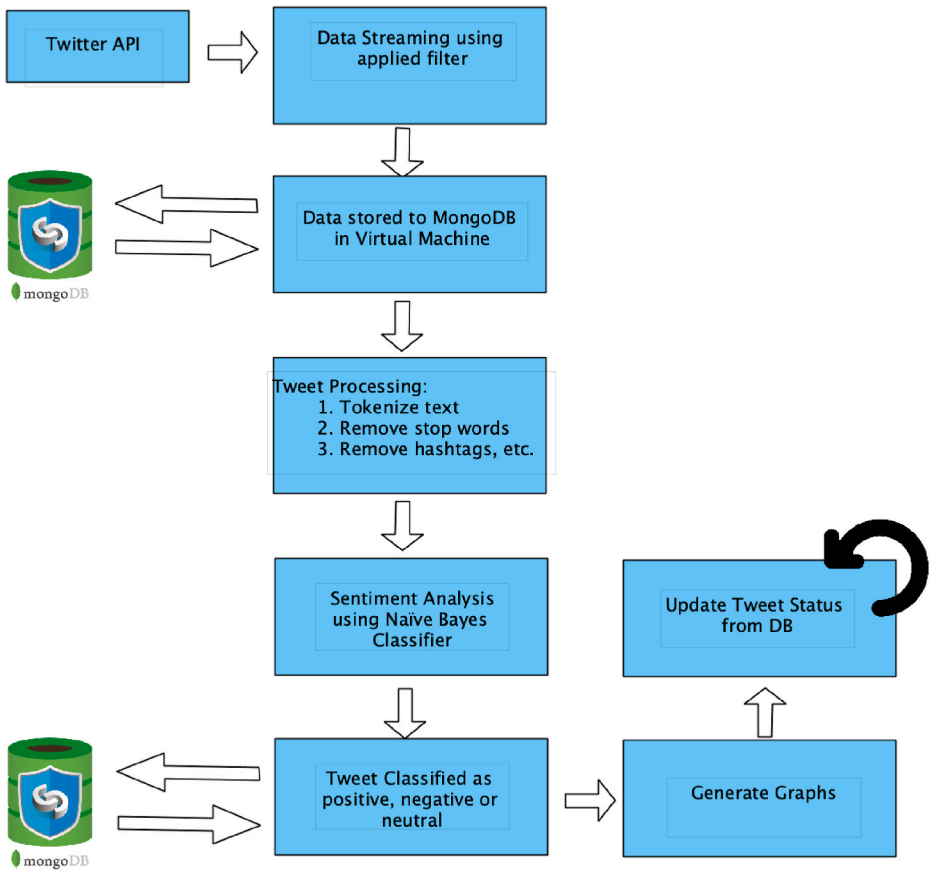


Fig. 1. Project architecture.

Table 1
Main Filters that we used to stream tweets.

Type	Filter applied
Minery Words	minery, open-pit mining, excavation, quarry, colliery
Hasthags	mining sustainability, mining purpose, responsible mining, future mining, women mining, dieselfree mining, coal mine, safetyfirst, environment, jobs, development, diversity, re-newable, safe, respect
CSR Words	CSR, ethic, ethics, ethical, responsible, responsibility, legal compliance, care, respect, reputation, social, socialgood, community, charity, philanthropy, volunteer, nonprofit, cul- ture, sustainable, sustainability, environment, renewable, climate, development, rehabilitation, stakeholder, code con- duct, eco, green, safety, safe, diversity, gender parity, trans-parency,labor practise, human rights

parameters summarized in Table 1. Each word of *Minery words* is combined with all the words of *CSR words* and with all the words of *Hashtags mining*. The result is a long list of word combinations that will ensure the tweets found are of the appropriate topic.

3.3. Algorithm

Algorithm 1 shows the system used to perform tweet streaming 24 h a day, every day of the week, to make sure we do not lose any tweets.

4. Classify algorithm

4.1. Introduction

In this section we present the method used to solve the problem of the classification of texts by their feeling at the document level. These documents are messages that have been published on the social network Twitter. A training corpus is required. Its examples were previously labeled one by one by hand with the category of feeling to which they belong (negative, positive or neutral)

Algorithm 1 Tweets Streaming Algorithm

```
procedure Stream Tweets(TwitterAPI, Filter, MongoDB)
    Initialize the connection to MongoDB
    Initialize the connection to the TwitterAPI through credentials
    Filter ← load list of words to filter while streaming
    while OnData ≠ Null do                          ▷ While we find a tweet:
        tweet = ConvertToDict(tw).lower()
        user = tweet['user']
        hashtags = tweet['hashtags']
        if tweet ∈ filter then
            if tweet ∉ MongoDB then
                MongoDB ← < tweet, user >
            end if
        else
            discard tweet
        end if
    end while
end procedure
```

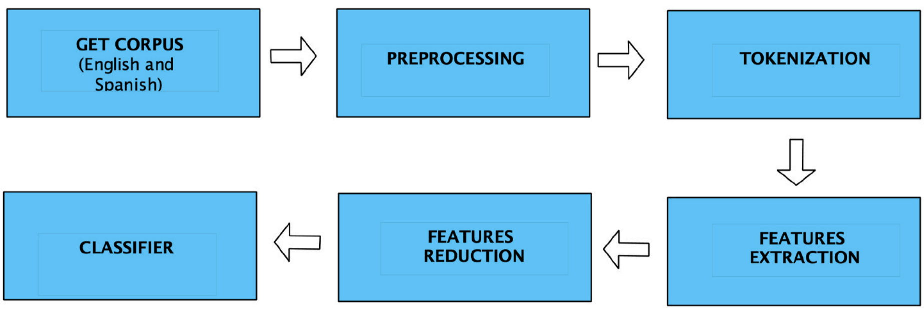


Fig. 2. Phases for training of the classifier.

4.2. Algorithm training process

As can be seen in Fig. 2, the usual construction process of a text classifier system based on machine learning consists of several sequential phases. First, it is necessary to prepare the data to train the algorithms. To do this, information must be cleaned and normalized in order to reduce or eliminate those data that may negatively influence the final result. Each of the texts undergoes then a process called tokenization, which divides them into smaller units or tokens and which are usually the words of the messages. From the tokens, the characteristics that represent the original messages are extracted and, optionally, a method can be applied to reduce their number. To finish, these characteristics are weighted according to the importance that they want to give them and with them the classifiers are trained.

Phase 1. Get corpus: In the case of tweets written in English, NLTK provides us with the corpus directly from its Python library, with more than 30,000 classified tweets, so we did not have to search for a corpus manually. In the case tweets in Spanish, we have obtained, as we mentioned previously, an XML file where we extract the manually classified tweets.

Phase 2. Preprocessing: In any method that makes use of automatic learning algorithms it is necessary to pre-treat the data with which they will be trained. The objective of this phase is to clean and normalize the information to prevent certain data from negatively influencing the final result. This question is crucial when we talk about messages extracted from social networks since it is very common to find messages with spelling mistakes, repetitions of characters, mixed uppercase and lowercase letters … For this research we have selected a set of simple rules to apply and which are common in the construction of this type of classifiers. The goal pursued is the normalization of messages, avoiding at all times that the changes applied cause the loss of the polarity of feeling.

a) Normalization of capitals and small letters: although for people it is easy to know that the words "dog" and "DOG" have exactly the same meaning, for machine learning algorithms they do not. In fact, they are treated as totally different words, without any relationship between them. To prevent this from happening, all messages are converted to their equivalent in lowercase letters.

b) Treatment of characters duplicity: in social networks it is common to repeat the same letters in words to give intensity to what is intended to express. For example, it is not the same to write “I am hungry” or “I am huuungry”. Although both sentences are conceptually equivalent, the second emphasizes the feeling of warmth by repeating the characters of the phrase. In our case, we do not need to know to what extent opinions are issued, only to know what polarity they have so it is not necessary to maintain these repetitions. Therefore, any sequence of three or more equal characters will be reduced to only two. For example, the previous case would be like “I am hungry”.

c) Elimination of line breaks: in some Twitter messages the text is written in different lines. These elements are also deleted so that the messages appear written in a single line.

d) Elimination of mentions and links: these two elements are very common in Twitter messages. The mentions are used to refer to other users of the social network by their name preceded by the symbol of the @ (for example, @RioTinto). Finally, it is possible to add to the tweets links to web pages to enrich the messages (for example: http://t.co/4diTkV2a).

Phase 3. Tokenization: Once the message normalization process is completed, the next phase is called tokenization. In this phase the texts are divided into smaller units called tokens, which normally correspond to the words of each text. This process can be as simple as separating the terms of the phrases by the white space and the punctuation characters or else consider that the grouping of certain symbols can contain some type of information that is useful to the classification process. This could be the case of emoticons, punctuation character sequences that are usually an indicator of the polarity of the feeling of the words they accompany. In our case, we will use a tokenizer that maintain the emoticons, mentions, hashtags and URLs as tokens.

Phase 4. Features Extraction: From the tokens obtained in the previous step, we will define how to represent with them the messages from which they come, thus creating the features. The usual task in text classification is to use the bag of words model (BoW) where each message is represented by its tokens without taking into account any specific order between them.

Phase 5. Features Reduction: This phase is optional and its objective is to decrease the number of characteristics of the corpus by eliminating certain tokens or converting them looking for the same way of representing them. There are two common techniques to carry out this task: stop words removal and stemming.

a) Elimination of stop words: there is a set of words that, although they are necessary to construct meaningful sentences, lack information that helps determine the polarity of the texts in which they are found. In Spanish and English, these words are prepositions, pronouns, conjunctions and the different forms of the verb “to be”, among others. By means of this technique, all the terms belonging to the stop words list will be eliminated from the model before the training of the algorithms.

b) Stemming: this is another method of morphological normalization.

In this case, a word is transformed to its root by means of the suppression of its suffixes and inflections. Following the previous example, the word “climate” would be converted to its root, “cli-mat”.

Phase 6. Classifier: Once the data is preprocessed, the Naive Classifier gets all this information and classifies the tweets according to their positivity.

4.3. Naive Bayes classifier

The Naive Bayes classifier is a probabilistic method for classification. It performs an approximate calculation of the probability that an example belongs to a class given the values of predictor variables. The simple Naive Bayes classifier is one of the most successful algorithms on many classification domains. In spite of its simplicity, it is shown to be competitive with other more complex approaches in several specific domains. This classifier learns from training data the conditional probability of each variable Xk given the class label c. Classification is then done by applying Bayes rule to compute the probability of C given the particular instance of X1,…, Xn,

$$P\ (C = c|X_1 = x1,\ \dots,\ Xn = xn)$$

Naıve Bayes is based on the assumption that variables are conditionally independent given the class. Therefore the posterior probability of the class variable is formulated as follows,

$$P\ (C = c|X_1 = x_1,\ \dots,\ X_n = x_n)P\ (C = c)P\ (X_k = xk|C = c)$$

This equation is highly appropriate for learning from data, since the probabilities pi = P(C = ci) and pik,r = P(Xk = xrk—C = ci) may be estimated from training data. The result of the classification is the class with highest posterior probability.

In the specific case of the classification of texts, the exclusive and exhaustive events are the different classes that can be assigned to a message, so that it is not possible to assign more than one simultaneously (excluding) and those classes are all types that exist (exhaustive) Naive Bayes algorithms are often referred to as “naive” because in their calculations the characteristics selected to represent the training examples are statistically independent and contribute equally in the classification process. In other words, and in the specific case of the classification of texts, it is considered that the words of the same message do not maintain any kind of relationship with each other and the position they have within the text to which they belong is indifferent.

One of the problems of supervised methods is the need to have a representative test set (previously labeled) to train the machine learning algorithms, that is, a corpus. As reviewed above, the classification performance can directly depends on the quality of features obtained from a training dataset. Thus, our choice for Naive Bayes algorithm appears to be reasonable in terms of accuracy for our dataset characteristics according to the results obtained from Choi and Lee (2017).

In the specific case of the classification of Twitter messages, although there are currently several resources in English, it is not so easy to find them in Spanish. The creation of this type of elements is often complicated due to the enormous cost in terms of time and effort necessary to complete it. This was the problem that the authors of the first work of sentiment analysis on Twitter found Go et al. (2009), since at that time Twitter was not a popular network and, therefore, there were still no available corpora. To solve this obstacle they designed an automated system that allowed creating a corpus following the ideas presented in Read (2005). Broadly speaking, Read (2005) stated that when a user added an emoticon to a certain text, this element was an indicator of the feeling of his written words. In other words: if a text was added to a text with a happy face: (the text would have a positive connotation. If instead the symbol was a sad face:), those words would be negative.

For the tests on this research project, we used an existing corpus in Spanish whose authorship belongs to the Workshop on Semantic Analysis (TASS[2]) of the Spanish Society for the Processing of Natural

[2] http://www.sepln.org/workshops/tass/.

Language (SEPLN[3]). This society annually organizes a competition in which different methods are presented for the classification of tweets in four and six categories.

At a technical level, it is worth mentioning that the models have been written in Python[4] and that specific libraries have been used for this type of development, such as NLTK,[5] which specializes in the processing of natural languages.

The Naive Bayes algorithms are based on probabilistic theories, which allow us to estimate the probability of an event based on the probability that another event will occur, on which the first one depends. Basically, it is about estimating the probability that a document belongs to a category. Such belonging depends on the possession of a series of characteristics, of which we know the probability that they appear in the documents that belong to the category in question. Naturally, these characteristics are the terms that make up the documents and both their probability of appearance in general, and the probability that they appear in the documents of a certain category, can be obtained from the training documents. For this, the frequencies of appearance in the collection used during the learning are used.

However, there is a very important problem that obliges to expand the mentioned considerations through more sophisticated versions of mentioned algorithm. When the statistics are calculated with respect to a class, the occurrences of the words in the training documents are counted. But if a new document containing a word (term) appears that has not appeared before, the probability of the term will be zero and, consequently, also the probability of belonging to the document that contains it will be zero. with respect to any class, leading to erroneous results. The number of errors will increase when small training collections are used, therefore if we use a small number of samples, it will be more difficult a priori to consider any possible word.

Algorithm 2 gives us a brief idea of the global structure we have applied to classify tweets according to their sentiment, as well as the deletion process in case the tweet does not belong to the Spanish or English language, which are the only languages we have analyzed. It is mainly a request in our database for tweets that have not yet been classified and apply the Naive Classifier for all of them.

5. Results

Next, the main results of the research conducted are discussed.

5.1. Word clouds

After the processing of all the data collected and the stemming of the text, the roots of the words were used to get a more accurate result of the filtered words. For example, the root *"climat"* refers to all the words that are written by those letters, such as *"climate"*. These words or combination of words (i.e. *"climate"*, *"climate-change"*, *"climate-debate"*) are the ones that Twitter users use when they talk about CSR in the mining industry, which indicates that when they hear about these terms those are the first ideas with which they relate the CSR. Fig. 3 gives a general vision of which words do people use when talking about topics related to CSR in the mining industry.

Algorithm 2 Classify Tweets with Naive Bayes

```
procedure Naive Bayes(CSRp, CSRrt, SearchEngine)
    Initialize the connection to MongoDB
    NotClassified ← load all tweets without classification from MongoDB
    DataSet ← Positive Tweets + Negative Tweets
    TrainingSet ← 0.8 * Positive Tweets + 0.8 *Negative Tweets
    StopwordsEnglish ← NLTK.stopwordsenglish
    StopwordsSpanish ← TASSstopwords
    NotClassified ← TweetsnotclassifiedfromDB
    while NotClassified ≠ Empty do
        if TweetLanguage = english then
            positivity ← ClassifyEnglishTweet
            MongoDB ← < positivity, tweet >
        end if
        if TweetLanguage = spanish then
            positivity ← ClassifySpanishTweet
            MongoDB ← < positivity, tweet >
        else
            delete tweet
        end if
    end while
end procedure
```

3 http://www.sepln.org/.
4 https://www.python.org/.
5 http://www.nltk.org/.

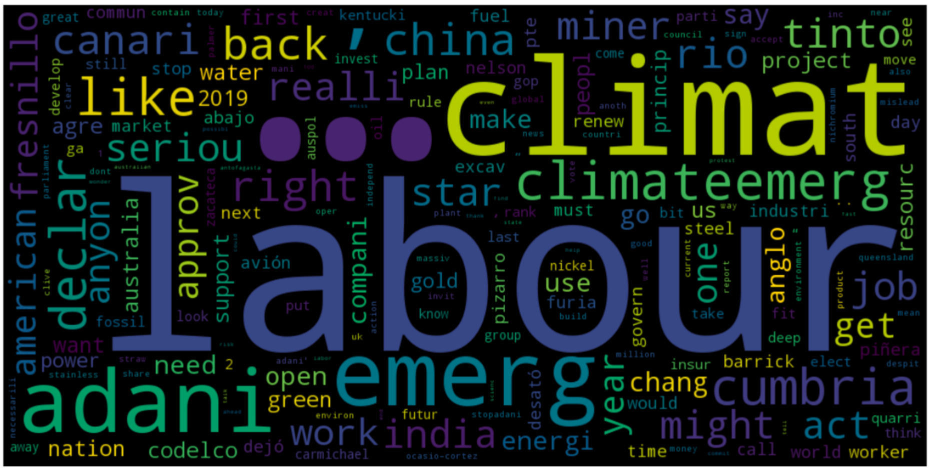


Fig. 3. Cloud of all words.

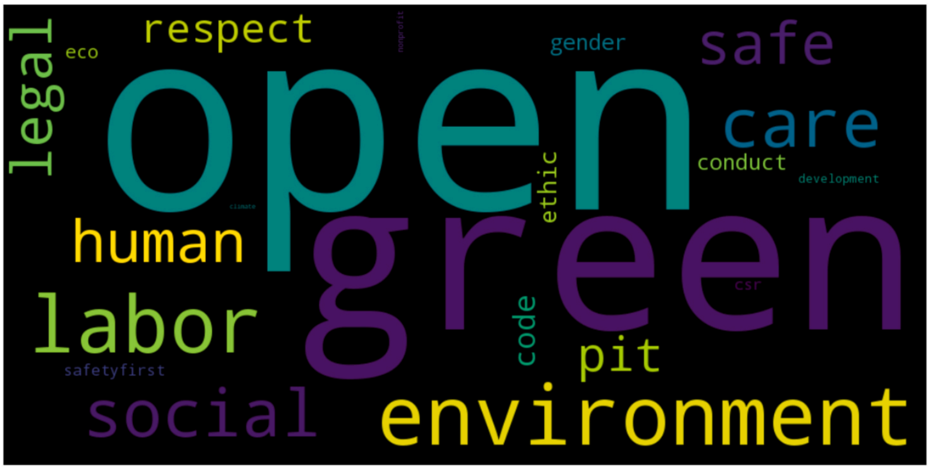


Fig. 4. Cloud of CSR words.

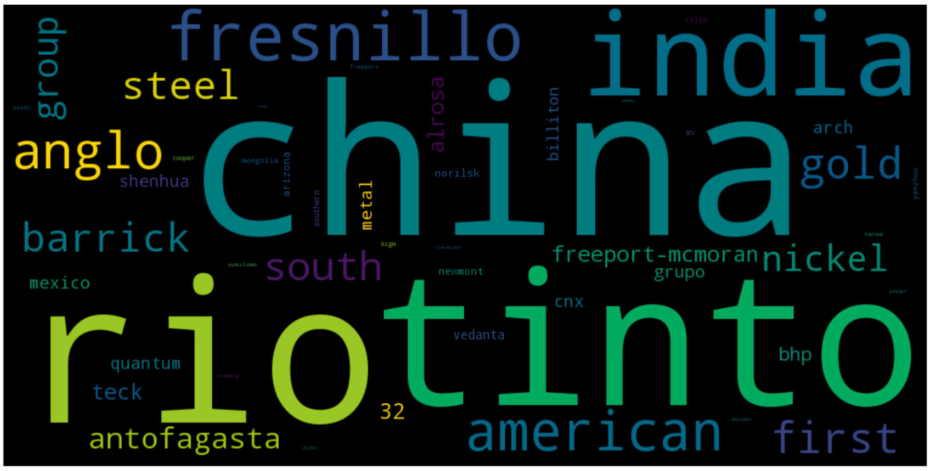


Fig. 5. Cloud of companies.

Fig. 4 gives a more detailed vision of results concerning specific CSR words. The most popular CSR terms are related to environmental issues, i.e. *"green"*, *"environment"*, *"open"* and *"eco"*, followed by words related to social issues including work or legal activity, such as *"labor"*, *"human"*, *"legal"*, *"gen-der"* and *"ethic"*. Another group of words may refer to both CSR categories: i.e. *"safe"*, *"respect"*, *"care"*, *"code"* and *"conduct"*.

These results are consistent with different studies published in the literature that reveal that one of the most visible impacts of mining is reflected in the environment Vintro et al. (2012) Jenkins and Yakovleva

(2006) and that CSR environmental actions (i.e. restoration plans, mining source reduction and control of energy consumption) are more widespread than social actions (i.e cooperation with non-governmental organizations and promotion of local communities) Vin-tro et al. (2014). This difference in the application of CSR practices as well as in the topics discussed on Twitter could be attributed to the high impact of the mining industry on its surroundings and on local communities, with morphological changes in the exploited areas, noise, dust, and surface and groundwater pollution Hilson and Murck (2000) Jenkins and Yakovleva (2006). In this sense, different tweets collected by social stakeholders (that is, tweets from anonymous personal accounts) argue that mining initiatives are still controversial when considering its impact on the land and advocate for more responsible environmental strategies:

> *"#stopadani India campaign must be started. Their coal mines are targetting multiple green zones. Don't destroy the land, our home for an unsustainable energy source. Ecosystem supported by the forests will never recover. water table,1000 yrs of growth #studentprotest #burning." (Tweet posted by a non- professional account)*

Moreover, the top environmental words found in the Twitter analysis fit with previous results that provide a description of good environmental practices adopted by mining companies. Vintro et al. (2014) observed that investing in processes for saving energy and natural resources, reducing green-house gas emissions, improving the safety of the work environment, restoring initial environmental conditions, and minimizing the impacts to the environment were some of the most prominent examples. Suppen et al. (2006) cite different cleaner production technologies. Sandoval et al. (2006) mention transparent policies and practices to ensure long-term economic, environmental and social well-being of local communities. And Kumar (2014) reviews sustainable mining practices including environmental impact assessment, geographic information system (GIS) maps, three-dimensional (3D) models, and the use of global positions system (GPS). On this respect, different tweets posted by mining companies and collected during the analysis inform about different initiatives and projects undertaken:

> *"Mining companies test and treat water on at least a monthly basis and Nova Scotia Environment is the regulator that ensures companies adhere to regulations. As a result, water is usually cleaner after it has been used on a mine/quarry than it was before". (Tweet posted by a professional/ official ac count)*

When referring to the top social words found in the Twitter analysis, results show they are also in line with previous researches. For instance, Amponsah-Tawiah and Mensah (2016) examine the relationship and impact of occupational health and safety on employees' organizational commitment in Ghana's mining industry. Jarosławska-Sobór (2015) analyzes the social potential of a coal mining company in terms of human capital and occupational safety. Maier et al. (2014) analyze the relationship between mining and poverty, human health and the environment. And Vin-tro et al. (2012) observed that codes of conduct and professional career programs were the most popular CSR practices linked to human resources management. Results from the Twitter analysis show that ethical concerns are expected to be integrated into operations and management, which fits with similar studies (Tate et al., 2010), and issues such as human rights, child labor and working conditions are big issues that mining stakeholders care about:

> *"I get hurt when I see Centre and the states are ignoring the voice of tribal coal mine workers in the country". (Tweet posted by a non-professional account) "Coal miners deserve better options. Coal miners deserve better pay. Coal miners deserve safer working conditions." Tweet posted by a non-professional account)*

Health-related issues are also discussed and seem to be of interest in the Twitter sphere, which is not strange considering mining has

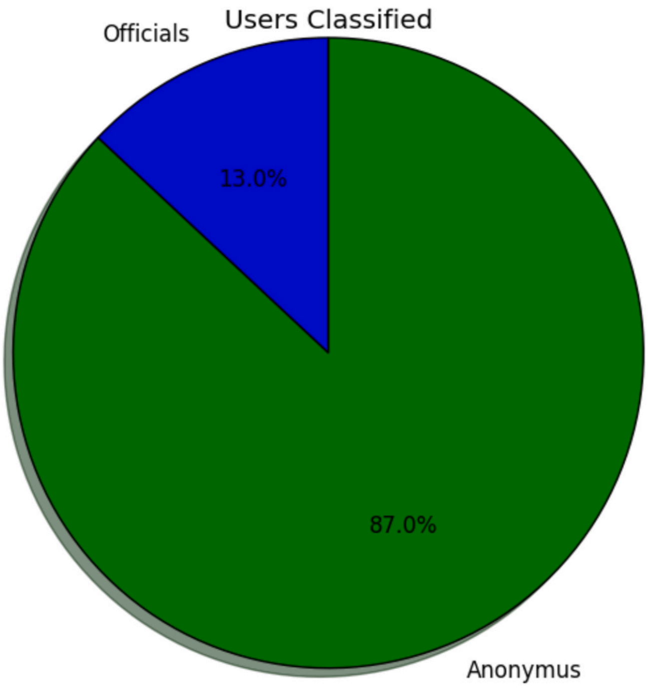


Fig. 6. Users classification by Official/Anonymous.

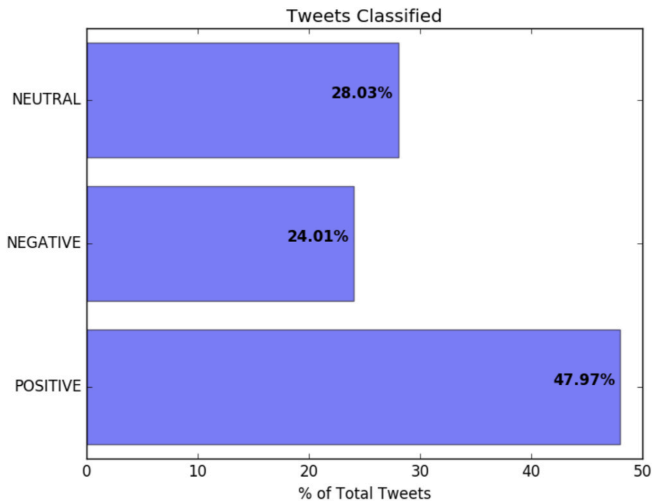


Fig. 7. Sentiment classification of the total Tweets.

traditionally been seen a dangerous activity due to its adverse social impacts when referring to health and safety problems Hilson and Murck (2000) Sánchez (1998):

> *"Van Howell leading the class through a workshop exercise in OSHA 7410: Managing Excavation Hazards @ Idaho Falls Safety Fest at the College of East- ern ID! The PNW OSHA Education Center was a founding partner of the Safety Fest of the Great Northwest starting in Boise in 2005! https://t.co/vf3e3VRpNa". (Tweet posted by a professional account)*

The economic aspect of CSR is also drawn in the results. Words such as "responsibility", "plan", "invest", "profit, "company", "charity", "project", or "corporate governance" detected in the tweets, reveal that CSR is a strategic decision to create sustainable and competitive businesses. The group of words referring to the economic dimension of CSR in the mining sector turned to be less popular on the Twitter debate.

Fig. 5 shows results of words concerning profiles of companies (i.e. mineral, country, company name) with the most Twitter activity.

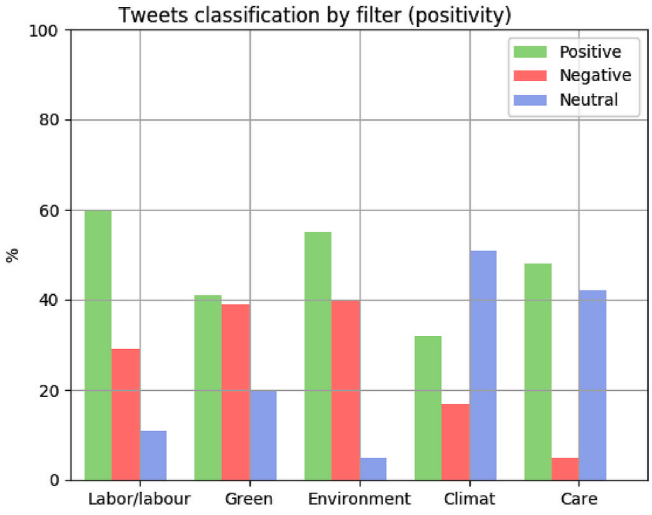


Fig. 8. Weighing with Näives Bayes of most used words.

Results reveal that companies use Twitter as a tool for promoting community-focused activities and philanthropic dimensions of CSR in order to reinforce a positive relationship with key stakeholders. CSR activities must be aligned with social norms and values to maintain and gain organizational legitimacy. Moreover, results show that interest in CSR is growing in developing countries Collier and Esteban (2007) and in countries with a bad reputation of environmental and health mining conditions Wei-ci and Chao (2011). In this line, two specific countries were popular and captured as top words: China and India. Also, different names of companies tagged are from South America. These results concur with the literature since an increasing number of general CSR studies focusing on emerging economies such as India Shirodkar et al. (2018) and specific CSR studies focusing on mining in emerging countries such as Ethiopia Woldeyohannes et al. (2018) can be found.

5.2. *Users classification: official and personal anonymous twitter accounts*

Fig. 6 depicts the percentage of tweets written by anonymous personal accounts (that is, public in general) and those written by official accounts (that is mining companies, mining associations and CEOs of these companies or associations). Results show that only a small amount of tweets (around 13%) are from official accounts. Official accounts generally have more followers and more influence on society than anonymous personal accounts. Therefore, mining companies and the mining sector in general should increase the adoption of new strategies to reach stakeholders satisfaction and transparent results reporting. According to a study published by Deloitte, 2016 based on tracking trends, the need for the mining sector to use social media is becoming increasingly important since it opens new opportunities for businesses to spread information and to engage stakeholders Cho et al. (2017) Kollat and Araujo (2018).

Some general studies about CSR disclosure on social media can be found in the literature. For example, Chae and Park (2018) conducted a study about CSR topics and trends in Twitter and concluded that Twitter is an important channel for organizations to communicate with stakeholders. The authors found that many organizations use the social media to promote their CSR practices. Specific studies referring to the mining sector are very scarce.

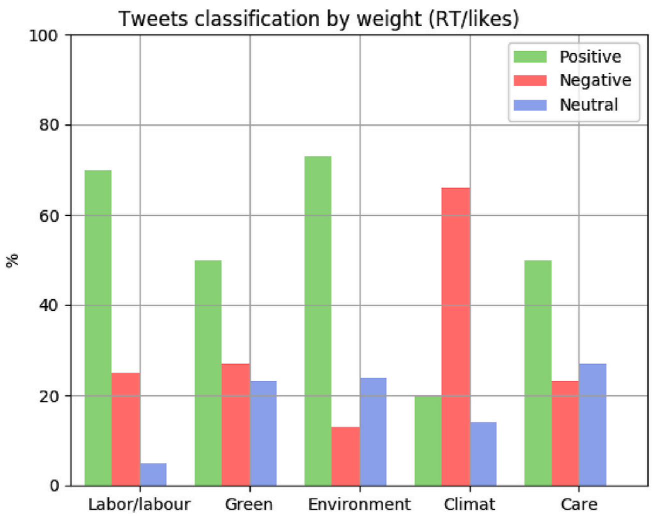


Fig. 9. Weighing with number of RT and Likes of most used words.

Table 2
Confusion matrix.

Confusion Matrix	Predicted NO	Predicted YES
Actual NO	78%	20%
Actual YES	22%	80%

5.3. *Sentiment analysis of tweets*

The results obtained with the user and Naive Bayes classifier (Fig. 7) show there is a slight tendency to classify the tweets as positive (around 48%). Approximately one in four tweets (25%) is considered negative, and neutral tweets account approximately for 28%.

Fig. 8 shows the results by CSR topic. We chose the 5 most popular words found in the database used for the investigation. There is not a concluding result that indicates which topics (environmental issues, social issues, …) are seen as more positive and which ones are seen as more negative.

Retweets and likes are indicators of the impact of tweets on society. In Fig. 9 we take into account retweets and likes when weighing the results. Results show more clearly which words are related to positive topics and which ones to negative. For instance, tweets with the word “environment” throw a positive sentiment while tweets with the word “climate” mainly have a negative mood. This last result fits with the study conducted by Cody et al. (2015) who found that tweets containing the word “climate” are less happy than all tweets, probably because Twitter users associate climate change with the in-crease in severity and frequency of certain environmental disasters and because the discussion on climate change on the Twitter sphere is dominated by climate change activists. On the other hand, tweets containing the word “environment” mainly refer to environmental and sustainable practices which are seen as valuable to fight against climate change and to improve the environmental behavior of companies.

5.4. *Confusion matrix Naive Bayes*

Table 2 shows the effectiveness of the developed Naive Bayes algorithm. Around 80% of the classifications have been made correctly, a fairly positive and effective number since most of these classifiers are around 75% effective.

6. Discussion

6.1. *Implications for theory*

In line with previous studies published in the literature Hilson and Murck (2000), Warhurst (2001b), Hilson and Nayee (2002), Dutta et al. (2012), we find empirical evidence that CSR in the mining sector is an increasingly studied topic by different researchers and is a core strategy of many companies.

While previous studies published in the literature have mainly used traditional methods for data collection such as reports, interviews or questionnaires, this investigation has introduced the use of big data, which opens up new opportunities for research and provides a valuable source of information. To take advantage, CSR researchers and practitioners need to be familiar with computational data collection techniques such as APIs and machine learning algorithms. Its performance strongly depends on the way of training a classification model and data properties. Twitter is the perfect data source due to maximum tweet length allowed as too long documents have noise features that hinders the sentiment analysis accuracies Choi and Lee (2017).

Our main contribution is the development of a database of tweets for the mining sector that will serve for future projects in this area, and an automatic Tweets classifier using the Naive Bayes algorithm. The results obtained here may have implications for understanding that CSR debate in the Twitter sphere is growing and that the open discussion is mainly centered on the impact of mining industries on the land and on ethical concerns referring to working conditions and occupational safety.

6.2. Implications for practice

Mining companies face a challenging competitive environment and must adapt and readapt strategies to respond positively to environmental and sustainable issues. They need to maintain a social license to operate and this is now directly linked to value perceived by stakeholders. Moreover, the digital age has turned the world and the competitive environment more transparent and companies should engage with different stakeholders in new ways. That is, the mining sector has traditionally had a business-to-business mentality, but now it must adopt a business-to-stakeholders approach.

As Lattemann and Stieglitz (2007) state, social media connects companies to their stakeholders and engage them to discuss opinions on current topics and trends. In this sense, the results of the study presented in this paper show that some mining companies have started to use social networks to provide a viable source of information and to create an environment for effective stakeholder dialogue, and different stakeholders play an active role in CSR discussions on social network sites such as Twitter.

For business, our study provides information about the main topics discussed by stakeholders which reveal the issues of concern and the perception they have about the initiatives undertaken by mining companies. This is a valuable source of information for the CSR strategy definition. The results also shed light on the need to cultivate long-term relationships with key stakeholders and to increase the use of social network sites such as Twitter for real company-stakeholder engagement and debate. It is clear that there are important benefits from knowing the main issues and concerns from stakeholders, therefore companies might redefine CSR role and priorities. This framework provides a platform for further research and application to other sectors.

For public administrations and governments, our findings seem to indicate the need for more strict environmental regulations and for actions to impulse the adoption of sustainable and ethical practices by mining companies, so that the negative impacts on the environment and society get reduced.

6.3. Limitations of the study and further research

Our research has some practical limitations, since it only considered the Tweets collected within a specific period of time and the Tweets written in English and Spanish. Moreover we developed a predefined list of specific initial parameters to filter the Tweets.

In this sense it would be interesting for further research to include more languages and more filters in order to extend the list of initial words to search within the Tweets. It would be also interesting to conduct similar studies analysing correlations between topics and words, to correlate topics and profiles of users (professional and non–professional accounts, geographical origin, …) and to use combinatorial analysis of the results with another social network to conduct a more detailed study of the CSR mining debate.

7. Conclusions

CSR communication has increased over the years, and many companies of different sectors communicate about economic, environmental and social issues in some form, and most of them engage in sustainability reporting (Frosten-son et al., 2012). Specifically, the mining industry faces pressure to adopt and communicate CSR practices.

Communicating effectively CSR activities benefits company-stakeholder engagement and develops positive attitudes from stakeholders towards the company. In this sense, some mining companies have started to use social networks to inform stakeholders and to promote their CSR practices, and now, different stakeholders play an active role in CSR discussions on social networks sites such as Twitter.

This study has aimed to understand what topics of CSR in the mining industry are discussed in the Twitter sphere. CSR appears to be an evolving construct in business and society Gupta (2015) and its dimensions and trends change over time. In this sense, the examination of big amounts of data obtained from social media may help in the aim of analyzing the CSR debate.

The results obtained in the investigation are consistent with different studies published in the literature that emphasize the open discussion about environmental issues (that is, environmental impacts of mining activities and environmental actions such as restoration plans or mining source reduction). The debate is mainly centered on the impact of mining industries on the land, and many stakeholders advocate for more environmental responsibility since they consider that mining activities are still controversial. Companies use Twitter to inform about different initiatives undertaken to be more sustainable and environmentally responsible. Social actions appear to be the second most discussed issue and the debate around it is centered in the improvement of ethical concerns considering human rights, child labor and working conditions. Occupational safety appears in many posts. The economic aspect of CSR seems to be the least commented topic in the Twitter sphere and is mainly centered on corporate governance issues.

The results show the CSR debate is increasingly growing in developing countries and in countries with a bad reputation of environmental and health mining conditions. On the other hand, the percentage of tweets posted by anonymous personal accounts (public in general) accounts for 87% while those written by official accounts (that is mining companies, mining associations and people related to these companies or associations) only reach 13%. These results reveal that the mining sector should improve the CSR disclosure and adopt a stakeholder engagement strategy grounded in the corporate stakeholder relationship perspective, the two-way symmetrical communication, and the dialogic theory of public relations Lim and Greenwood (2017). Social network sites such as Twitter can lead to positive outcomes for companies since they are not only a way to advertise what the company does with regards to CSR, but also to receive input from other for their CSR activities. Different authors defend that social media make corporation communications more strategic, more interactive and more socially responsible Grunig (2009).

Results also reveal a slight tendency to positive tweets (48%) in front of negative tweets (25%) and neutral tweets (28%). There is not a concluding result that indicates whether environmental issues, social issues, or economic issues are seen as more positive and which ones are seen as more negative. For instance, when considering environmental issues, tweets with the word “environment” have a positive mood and are aligned with the value of environmental and sustainable practices undertaken by mining companies. On the other hand, tweets with the word “climate” are less happy and associate climate change with the increase in severity and frequency of certain environmental disasters.

Author statement

Adrià Pons (Ph.D. Student in Business Administration)

Adrià is a PhD student and this work is part of his thesis. He participated in the analysis of data and in writing the manuscript. He also revised and edited the document.

Carla Vintrò (Ph.D. in Natural Resources and Environment)

Carla was responsible of the conceptualization and formal analysis. She supervised the analysis and interpretation of data and participated in writing the manuscript.

Josep Rius (Ph.D. in Computer Science)

Josep designed the methodology and the software used to connect with twitter and collect the data. Together with Carla, he also supervised Adrià in all phases and participate in writing the manuscript.

Jordi Vilaplana (Ph.D. in Computer Science)

Jordi was the main developer of the software tool used to collect the analyzed data. Also, helped creating scripts to manage and get the results from this data. He also participated in writing the manuscript.

Acknowledgments

This work was supported by the Ministerio de Economía y Competitividad under contract TIN2017-84553-C2-2-R. Also, the authors are members of the research group 2017-SGR363, funded by the Generalitat de Catalunya. Finally, we thank Victor Ferrer for his contribution.

References

Amponsah-Tawiah, K., Mensah, J., 2016. Occupational health and safety and organizational commitment: evidence from the ghanaian mining industry. Saf. Health Work 7, 225–230. https://doi.org/10.1016/j.shaw.2016.01.002.

Andriof, J., Waddock, S., 2002. Unfolding stakeholder engagement. In: Unfolding Stakeholder Thinking: Theory, Responsibility and Engagement, pp. 19–42.

Barnes, N., M, E., 2010. The fortune 500 and social media: a longitudinal study of blogging and twitter usage by America's largest companies. In: UMass Dartmouth Center for Marketing Research, vol. 6, p. 2011.

boyd, d., Golder, S., Lotan, G., 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, pp. 1–10. https://doi.org/10.1109/HICSS.2010.412.

Carroll, A., 1991. The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders. Bus. Horiz. 34, 39–48. https://doi.org/10.1016/0007-6813(91)90005-G.

Chae, B., Park, E., 2018. Corporate social responsibility (csr): a survey of topics and trends using twitter data and topic modeling. Sustainability 10, 2231. https://doi.org/10.3390/su10072231.

Cho, M., Furey, L.D., Mohr, T., 2017. Communicating corporate social respon- sibility on social media: strategies, stakeholders, and public engagement on corporate facebook. Bus. Prof. Commun. Q. 80, 52–69. https://doi.org/10.1177/2329490616663708.

Choi, Y., Lee, H., 2017. Data properties and the performance of sentiment clas- sification for electronic commerce applications. Inf. Syst. Front 19, 993–1012.

Claver Cortés, E., Molina-Azorin, J., Dolores López Gomer, M., Zaragoza-Sáez, P., 2004. La integraci'on del capital medioambiental en el capital intelectual de la empresa. Revista de economía y empresa, pp. 11–28. ISSN 0213-2834 21.

Cody, E.M., Reagan, A.J., Mitchell, L., Dodds, P.S., Danforth, C.M., 2015. Climate change sentiment on twitter: an unsolicited public opinion poll. PloS One 10, 1–18. https://doi.org/10.1371/journal.pone.0136092.

Colleoni, E., 2013. CSR communication strategies for organizational legitimacy in social media. Corp. Commun. Int. J. 18, 228–248. https://doi.org/10.1108/13563281311319508.

Collier, J., Esteban, R., 2007. Corporate social responsibility and employee commitment. Bus. Ethics Eur. Rev. 16, 19–33. https://doi.org/10.1111/j.1467-8608.2006.00466.x.

Cortes, C., Jackel, L.D., Chiang, W.P., 1995. Limits on learning machine accu- racy imposed by data quality. In: Advances in Neural Information Processing Systems, pp. 239–246.

Crane, L., Self, R., 2014. Big data analytics: a threat or an opportunity for knowledge management?. In: Lecture Notes in Business Information Processing 185, pp. 25–34. https://doi.org/10.1007/978-3-319-08618-7_3.

Dahlsrud, A., 2008. How corporate social responsibility is defined: an anal- ysis of 37 definitions. Corp. Soc. Responsib. Environ. Manag. 15, 1–13. https://doi.org/10.1002/csr.132.

Dellarocas, C., 2003. The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag. Sci. 49, 1275–1444. https://doi.org/10.2139/ssrn.393042.

Deloitte, 2016. The top 10 issues mining companies will face in the coming year. Track. Trends 50.

Dutta, S., Lawson, R., Marcinko, D., 2012. Paradigms for sustainable develop- ment: implications of management theory. Corp. Soc. Responsib. Environ. Manag. 19, 1–10. https://doi.org/10.1002/csr.259.

Escanciano, C., Fernández Muñiz, B., Sánchez, A., 2010. Organizaci'on de la actividad preventiva y gesti'on de la seguridad y salud laboral en la miner'ıa española: experiencia de las empresas certificadas iso 9001. Dir. Organ. 40, 86–98.

Fonseca, A., 2010. How credible are mining corporations' sustainability reports? a critical analysis of external assurance under the requirements of the inter- national council on mining and metals. Corp. Soc. Responsib. Environ. Manag. 17, 355–370. https://doi.org/10.1002/csr.230.

Go, A., Bhayani, R., Huang, L., 2009. Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report, Stanford 1, 2009.

Grunig, J., 2009. Paradigms of global public relations in an age of digitalisation. PRism 6, 1–19.

Gupta, S., 2015. CSR - Summarizing Evolution of the Concept. 4th Annual International Commerce Convention, vol. 4. Dept. of Commerce, Delhi School of Economics, University of Delhi, India, p. 18. https://doi.org/10.13140/RG.2.1.1053.4004.

Hamann, R., 2004. Corporate social responsibility, partnerships, and institu- tional change: the case of mining companies in South Africa. Nat. Resour. Forum 28, 278–290. https://doi.org/10.1111/j.1477-8947.2004.00101.x.

Hassan, A., Ibrahim, E., 2012. Corporate environmental information disclosure: factors influencing companies' success in attaining environmental awards. Corp. Soc. Responsib. Environ. Manag. 19, 32–46. https://doi.org/10.1002/csr.278.

Hilson, G., Murck, B., 2000. Sustainable development in the mining industry: clarifying the corporate perspective. Resour. Pol. 26, 227–238. https://doi.org/10.1016/S0301-4207(00)00041-6.

Hilson, G., Nayee, V., 2002. Environmental management system implementa- tion in the mining industry: a key to achieving cleaner production. Int. J. Miner. Process. 64, 19–41. https://doi.org/10.1016/S0301-7516(01)00071-0.

Hwang, S., 2013. The effect of twitter use on politicians' credibility and attitudes toward politicians. J. Publ. Relat. Res. 25, 246–258. https://doi.org/10.1080/1062726X.2013.788445.

Jarosławska-Sobór, S., 2015. Social potential growth of a mining company on the basis of human capital and occupational safety. J. Sustain. Min. 14, 195–202. https://doi.org/10.1016/j.jsm.2016.02.002.

Jenkins, H., Yakovleva, N., 2006. Corporate social responsibility in the mining industry: exploring trends in social and environmental disclosure. J. Clean. Prod. 14, 271–284. https://doi.org/10.1016/j.jclepro.2004.10.004.

Kelleher, T., 2007. Public Relations Online: Lasting Concepts for Changing Media. SAGE Publications, Inc. https://doi.org/10.4135/9781452225876.

Kira, K., Rendell, L.A., 1992. A practical approach to feature selection. In: Machine Learning Proceedings 1992. Elsevier, pp. 249–256.

Kollat, J., Araujo, T., 2018. Communicating effectively about csr on twitter: the power of engaging strategies and storytelling elements. Internet Res. 28, 419–431. https://doi.org/10.1108/IntR-04-2017-0172.

Kühl, N., Goutier, M., Ensslen, A., Jochem, P., 2019. Literature vs. twitter: em- pirical insights on customer needs in e-mobility. J. Clean. Prod. 213, 508–520. https://doi.org/10.1016/j.jclepro.2018.12.003.

Kumar, N.P., 2014. Review on sustainable mining practices. Int. Res. J. Earth Sci. 2, 26–29.

Lattemann, C., Stieglitz, S., 2007. Online Communities for Customer Relationship Management on Financial Stock Markets - a Case Study from a Project at the Berlin Stock Exchange, p. 76.

Lee, S.Y., Carroll, C., 2011. The emergence, variation, and evolution of corporate social responsibility in the public sphere, 1980–2004: the expo- sure of firms to public debate. J. Bus. Ethics 104, 115–131. https://doi.org/10.1007/s10551-011-0893-y.

Lee, K., Oh, W.Y., Kim, N., 2013. Social media for socially responsible firms: analysis of fortune 500's twitter profiles and their csr/csir ratings. J. Bus. Ethics 118, 791–806. https://doi.org/10.1007/s10551-013-1961-2.

Lim, J.S., Greenwood, C.A., 2017. Communicating corporate social responsi- bility (csr): stakeholder responsiveness and engagement strategy to achieve csr goals. Publ. Relat. Rev. 43, 768–776. https://doi.org/10.1016/j.pubrev.2017.06.007.

Maier, R., Díaz-Barriga, F., Field, J., Hopkins, J., Klein, B., Poulton, M., 2014. Socially responsible mining: the relationship between mining and poverty, human health and the environment. Rev. Environ. Health 29, 791–806. https://doi.org/10.1515/reveh-2014-0022.

Marwick, A., Boyd, D., 2011. To see and be seen: celebrity practice on twit- ter. Convergence. Int. J. Res. New Media Technol. 17, 139–158. https://doi.org/10.1177/1354856510394539.

Mclaren, S., Wehrmeyer, W., Argust, P.W., Graham, S., Robertson, J., 1999. Sustainability and the primary extraction industries: theories and practice. Resour. Pol. 25, 277–286. https://doi.org/10.1016/S0301-4207(00)00003-9.

Mendes, L., Rodrigues, M., 2018. Mapping of the literature on social respon- sibility in the mining industry: a systematic literature review. J. Clean. Prod. 181, 88–101. https://doi.org/10.1016/j.jclepro.2018.01.163.

Mitchell, T.M., 1999. Machine learning and data mining. Commun. ACM 42, 30–36.

Mudd, G.M., 2010. The environmental sustainability of mining in Australia: key mega-trends and looming constraints. Resour. Pol. 35, 98–115. https://doi.org/10.1016/j.resourpol.2009.12.001.

Newbold, J., 2006. Chile's environmental momentum: iso 14001 and the large- scale mining industry – case studies from the state and private sector. J. Clean. Prod. 14, 248–261. https://doi.org/10.1016/j.jclepro.2004.05.010.

Palazzo, G., Scherer, A., 2006. Corporate legitimacy as deliberation: a com- municative framework. J. Bus. Ethics 66, 71–88. https://doi.org/10.1007/s10551-006-9044-2.

Raufflet, E., Barin Cruz, L., Brès, L., 2014. An assessment of corporate social responsibility practices in the mining and oil and gas industries. J. Clean. Prod. 84, 256–270. https://doi.org/10.1016/j.jclepro.2014.01.077.
Read, J., 2005. Using emoticons to reduce dependency in machine learning tech- niques for sentiment classification. Assoc. Comput. Ling. 43–48.
Rodrigues da Silva Enríquez, M.A., Drummond, J., 2007. Social-environmental certification: sustainable development and competitiveness in the mineral industry of the brazilian amazon. Nat. Resour. Forum 31, 71–86. https://doi.org/10.1111/j.1477-8947.2007.00127.x.
Rybalko, S., Seltzer, T., 2010. Dialogic communication in 140 characters or less: how fortune 500 companies engage stakeholders using twitter. Publ. Relat. Rev. 36, 336–341. https://doi.org/10.1016/j.pubrev.2010.08.004.
Sánchez, L., 1998. Industry response to the challenge of sustainability: the case of the canadian nonferrous mining sector. Environ. Manag. 22, 521–531.
Sandoval, M.C., Veiga, M.M., Hinton, J., Sandner, S., 2006. Application of sustainable development concepts to an alluvial mineral extraction project in lower caroni river, Venezuela. J. Clean. Prod. 14, 415–426. https://doi.org/10.1016/j.jclepro.2004.10.007. improving Environmental, Economic and Ethical Performance in the Mining Industry. Part 1. Environmental Management and Sustainable Development.
Sanmiquel, L., Rossell Garriga, J., Vintro, C., 2015. Study of Spanish mining accidents using data mining techniques. Saf. Sci. 75, 49–55. https://doi.org/10.1016/j.ssci.2015.01.016.
Shirodkar, V., Beddewela, E., Richter, U.H., 2018. Firm-level determinants of political csr in emerging economies: evidence from India. J. Bus. Ethics 148, 673–688. https://doi.org/10.1007/s10551-016-3022-0.
Sinding, K., 2009. Environmental impact assessment and management in the mining industry. Nat. Resour. Forum 23, 57–63. https://doi.org/10.1111/j.1477-8947.1999.tb00238.x.
Smith, B.G., 2010. Socially distributing public relations: twitter, Haiti, and interactivity in social media. Publ. Relat. Rev. 36, 329–335. https://doi.org/10.1016/j.pubrev.2010.08.005.
Stelzner, M., 2012. 2012 Social Media Marketing Industry Report, How Mar- Keters Are Using Social Media to Grow Their Businesses. Social Media Examiner. Technical Report.
Suppen, N., Carranza, M., Huerta, M., Hernández, M.A., 2006. Environ- mental management and life cycle approaches in the mexican mining in- dustry. J. Clean. Prod. 14, 1101–1115. https://doi.org/10.1016/j.jclepro.2004.12.020.
Tate, W.L., Ellram, L.M., Kirchoff, J.F., 2010. Corporate social re-sponsibility reports: a thematic analysis related to supply chain management. J. Supply Chain Manag. 46, 19–44. https://doi.org/10.1111/j.1745-493X.2009.03184.x.
Tredinnick, L., 2006. Web 2.0 and business: a pointer to the intranets of the future? Bus. Inf. Rev. 23, 228–234. https://doi.org/10.1177/0266382106072239.
Türkel, S., Akan, A., 2015. Corporate Social Responsibility (CSR) Communi- Cation. A Turkish Industry Example, London, pp. 151–174. https://doi.org/10.1057/9781137388551_7.
Van Zyl, D., Lohry, J., Reid, R., 2007. Evaluation of resource management plans in Nevada using seven questions to sustainability. In: Proceedings of the 3rd International Conference on Sustainable Development Indicators in the Mineral Industries, pp. 403–410.
Vintro, C., Fortuny-Santos, J., Sanmiquel, L., Freijo, M., Edo, J., 2012. Is corpo- rate social responsibility possible in the mining sector? Evidence from Catalan companies. Resour. Pol. 37, 118–125. https://doi.org/10.1016/j.resourpol.2011.10.003.
Vintro, C., Sanmiquel, L., Freijo, M., 2014. Environmental sustainability in the mining sector:evidence from Catalan companies. J. Clean. Prod. 84, 155–163. https://doi.org/10.1016/j.jclepro.2013.12.069.
Warhurst, A., 2001a. Corporate citizenship and corporate social investment drivers of trisector partnerships. J. Comput. Chem. 1, 57–73.
Warhurst, A., 2001b. Corporate citizenship and corporate social investment drivers of trisector partnerships. J. Comput. Chem. 1, 57–73.
Warhurst, A., Noronha, L., 2009. Corporate strategy and viable future land use: planning for closure from the outset of mining. Nat. Resour. Forum 24, 153–164. https://doi.org/10.1111/j.1477-8947.2000.tb00939.x.
Waters, R., Jamal, J., 2011. Tweet, tweet, tweet: a content analysis of nonprofit organizations' twitter updates. Publ. Relat. Rev. 37, 321–324. https://doi.org/10.1016/j.pubrev.2011.03.002.
Wei-ci, G., Chao, W., 2011. Comparative study on coal mine safety between China and the us from a safety sociology perspective. Procedia Eng. 26, 2003–2011. https://doi.org/10.1016/j.proeng.2011.11.2397 iSMSSE2011.
Wheeler, D., Fabig, H., Boele, R., 2002. Paradoxes and dilemmas for stake- holder responsive firms in the extractive sector: lessons from the case of shell and the ogoni. J. Bus. Ethics 39, 297–318. https://doi.org/10.1023/A:1016542207069.
Woldeyohannes, D.Y., Salem, H., Ayongaba, B., Veljkovic, Z., 2018. Mining sector challenges in developing countries, tigray, Ethiopia and inspirational success stories from Australia. Int. J. Min. Miner. Eng. 9, 321–367. https://doi.org/10.1504/IJMME.2018.10018510.
Xiang, Z., Gretzel, U., 2010. Role of social media in online travel information search. Tourism Manag. 31, 179–188. https://doi.org/10.1016/j.tourman.2009.02.016.
Zhang, B., Marita, V., Veijalainen, J., Wang, S., Kotkov, D., 2016. The issue arena of a corporate social responsibility crisis-the volkswagen case in twitter. Stud. Media Commun. 4, 32–43.
Zhu, Q., Geng, Y., Fujita, T., Hashimoto, S., 2010. Green supply chain management in leading manufacturers: case studies in Japanese large companies. Manag. Res. Rev. 33, 380–392. https://doi.org/10.1108/01409171011030471.

CSR in leather industries

Adrià Pons[a], Josep Rius[a,1], Carla Vintró[a], Agapito Gallart[a]

[a] *Dept. of Business Administration, University of Lleida, Jaume II 73, 25001 LLeida, Spain.*

Abstract

Leather industry is facing the consequences of their historic lack of communication and now are under an increasing pressure to communicate their position and policies with regards to corporate social responsibility (CSR), informing consumers about the corporations' good intentions and actions to recover clarity and be able to affirm without doubts that responsibly manufactured leather is sustainable. Under this context, this paper aims to examine CSR communication in the leather sector on Twitter identifying the main topics of CSR and the main participants in the creation of content and thus contribute to the development of CSR strategies for the leather industry. A software framework has been implemented in order to manage big text data in the study of CSR and sentiment computational content analysis to classify the tweets as positive, negative or neutral. Data collection period was from January 2021 to July 2021 and it resulted in 90.000 Twitter posts. The results show that the general public associates leather with the environment, sustainability and fashion. The general opinion about the leather industry's CSR is rather neutral. Nonetheless, negative opinions occupy a close second place, and positive opinions are rare. Lastly, alternative bio-materials sneak in the industry reshaping the opinions to more positive ones.

Keywords: CSR; Leather industry; Twitter; Big Data; Sentiment Analysis

[1] CONTACT Josep Rius. Email: josep.riustorrento@udl.cat. Phone: +34 973 702 711

1. Introduction

The relevance of Corporate social responsibility (hereinafter CSR) in business management is behind the growing importance of CSR communication Moreno and Capriotti (2009), particularly for those sectors historically considered by the society as a threat to the natural surroundings, with environmental effects. Some might think that these industries are engaging CSR as a 'window dressing' to legitimize questionable business and deceive stakeholders, but what if society is still including wrongly one sector under the same umbrella? This is the case of leather industry which is trying to make a new start and desist from the historical burden. Considered a dirty job, nasty and without any social prestige from a hard and difficult sector which starts from an awkward but inevitable fact; the death of the animal. The industry has been traditionally located in the worst areas (outside the walls), physically cornered, expelled from the city. These traits influenced its industry personality making them reserved and unused to communication. In addition, the fact of commonly being a family business, rather small and very close to each other, has fostered this suspicious dynamic. This lack of communication from the industry geared towards its customers, consumers and the community at large has led to consider it in the past as one of the most polluting industries in the world due to the usage of significant quantities of chemicals, energy and water as well as the potential release of harmful substances during leather production process Kanagaraj et al. (2015) Sathish et al. (2016). Based on our 21st century knowledge, we acknowledge that leather making in history would not be categorised as sustainable today. Processes and chemicals that were used in some periods of history were thought to be safe at the time. As scientific knowledge improved some were identified as problematic or potentially dangerous. There are many similar instances in other industries such as cosmetics and medicines where lead, arsenic. cadmium and other poisonous materials were used extensively in industry before we understood their dangers, while "Progress" was seen as important and no one had properly realised the environmental damage it could do as industry grew bigger.

The consequences of this may be dire, particularly for sustainability which has become a “weasel” word thrown carelessly into documents, publicity material and conversations often with only limited interest in the truth.

Nowadays the industry do not avoid self-criticism and assume bad practices of the past Redwood (2018) investing resources mainly in the environmental sustainability aspect Omoloso et al. (2021). They are becoming proud of their industrial transformation forced by the strict environmental regulations together with the professionalization and modernization of new generations entering to the business. This led to significant advances and it is safe to say that the sector has been improved in all its areas; wastewater control, animal welfare, reduction of chemicals, energy efficiency, etc, however there’s a feeling of fear that its evolution has not turned into a change of the industry perception.

Leather sector plays a fundamental role upcycling an unavoidable waste from the food industry, to produce a versatile, durable, unique material, ideal for the circular economy that the world must move towards.. Leather making is a very old traditional process that serves social needs, employs many skilled/unskilled people while being one of the highest contributors of global economy Kanagaraj et al. (2015). But more important than this, the world needs materials that are sustainable, renewable, recyclable, biodegradable, and most importantly, do not add to the burden of atmospheric carbon. Natural fibres, such as leather, are part of the biogenic carbon cycle and as such are comprised of carbon that has been in the atmosphere for a millennia. These readily available raw materials, when ethically and properly produced, are an important replacement for fossil fuels, reducing the need for its extraction and retaining more carbon in the earth.

It is clear then, leather tanning industry need to focus special attention to their CSR communication in order to change stakeholder’s perspective. Research on CSR may facilitate a strategy to grasp stakeholder’s interests, enhance dialogue, meet their demands and expectations as well as be transparent when reporting back their actions, therefore generate real value for their different interest groups. According with Pons et al. (2021) it may also be worth to measure the impact of the actions undertaken by the companies together with interpretations and feelings being generated towards each interest group, and thus be able to redefine priorities and responsibilities within the organization from a triple perspective: economic, social and environmental.

The arrival of big data has shown an improvement in the identification and analysis of perceptions and sentiments from stakeholders which are relevant for all sectors, but even more for those seen as a threat such as Leather tanning. Now, debates have mainly shifted to social media where either internal and external stakeholders discuss and criticize in relation to actions taken by companies and its consequences. Under this context, this paper aims to examine CSR communication in the leather sector on Twitter and identify the main topics of CSR and the main participants in the creation of content. In other words, the results obtained will contribute to the development of CSR strategies for the leather industry and similarly other studies may be conducted to redefine CSR roles and priorities. The study is sustained in how organizational research focused on social media may translate stakeholder engagement into organizational goals and create the basis of effective strategy development. Twitter was chosen since it is one of the most commonly used social network sites Alharbi and de Doncker (2019) where opinions or expressions of sentiment about organizations, products, events, and people have proven extremely useful for social studies Thelwall et al. (2011). The data collection period was from January 2021 to July 2021 and it resulted in 90.000 Twitter posts.

In the next section, the paper details the architecture of the framework, describing its functionality and the algorithms implemented are then introduced. Next, the results obtained are presented. Discussions and conclusions follow.

2. Methodology

This paper aims to provide an interactive automatic system which predicts the sentiment of the review/tweets of the people posted in twitter using an automatic micro-service completely independent hosted on a server designed to be integrated inside a back-end ecosystem. Due to time limitations, only one social network source of information is processed, notwithstanding the service has a structure that allows the integration of more than one incoming source of information.

The work can be broadly separated into two processes, the first receives data and stores it, the second calculates and writes in graphics the results. In the interest of clarity, hereinafter referred to as `hub` and `analyzer` respectively. This method requires communication between them in order to share data collected. This is schematically shown in figure 1.

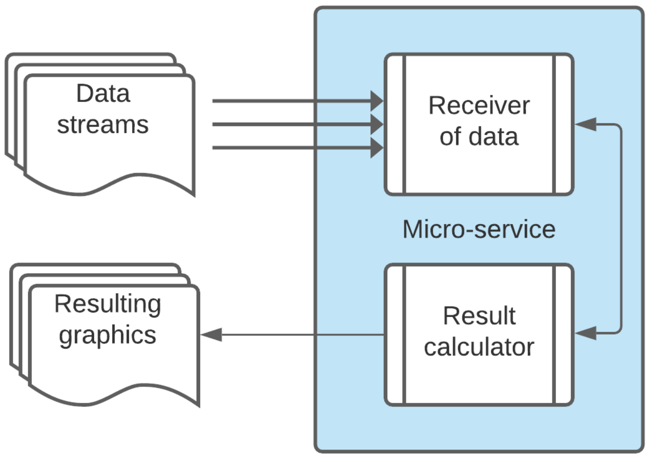


Figure 1: Basic schema of the micro-service

2.1. Twitter Api

The `Twitter Streaming API`[2] is used to collect streaming Tweets and notifications from Twitter in real time. However it requires a high performance, persistent, always in the connection between the server and Twitter. It searches

[2] `https://developer.twitter.com/en.html`

for hashtags, keywords and geographic bounding boxes simultaneously. The filter API helps for searching and delivers the continuous stream of Tweets which matches the filter tag. The work of the streaming implementation is to record the incoming events as quickly as possible and process them in the background. However, this suffers from limitations due to we only use the free version of the API which means it gives limited information about each bit of data. Consequently, it is required to call back using the identifier of each tweet to get the whole information.

2.2. Streaming tweets

Figure 2 shows the structure used for the streaming of tweets. The hub is the responsible of receiving and storing live data from Twitter into the database, however its design has been thought for a potential extension and comply with the requirements and necessities of adding multiple social networks (hence its name).

The sub-processes named SNX [3] are responsible for receiving data streams from SN (in figure 2 labeled as *Stream SNX*[3]). We have proposed a two levels approach for a better performance, SNXs may call other sub-processes[4] to handle multiple streams from the same source.

As the figure 2 reflect, the sub-process *SNX* can only handle up to one stream. This limitation prevents overload work for any *SNX* and keeps the structure of the code clear. Although, different *SNX* imply different algorithms for gathering data, they have to save the data at the same location.

The technologies used for that purpose have been the programming language Python[5], MongoDB, Mongo-Express and Alpine. Python was chosen as the programming language for its simplicity, performance, and flexibility. Mon-

[3] Where X is a number
[4] Labeled as SNX.Y where X refer to their parent process
[5] `https://www.python.org/`.

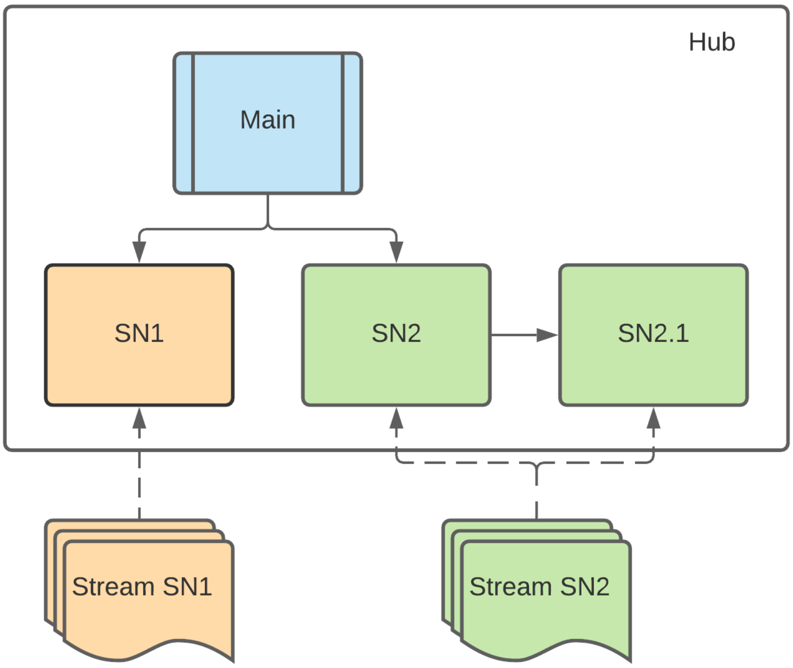


Figure 2: Hub's structure

goDB is a document-oriented database with dynamic schema. This allows to offer high performance and facilitates the development of applications. In turn, it prevents from having Joins and Transactions, something very common in relational databases. It should be noted that MongoDB stores documents in JSON format and its structures Chodorow (2013). Therefore, Mongo-Express was needed as it is a web-based MongoDB interface for administration purposes. Not only allows to manage databases remotely in a user-friendly manner, but also helps to monitor the service.

This approach requires an optimal solution to backup data periodically given that all the data gathered is a critical asset. For this purpose we used Alpine, which is a lightweight version of Linux appropriate for backups. Its container runs a script that copies all data from the corresponding social network. The script is stored in the crontab and executed daily Kerrisk. For security reasons, we suggest to register a log entry every time a backup is registered successfully.

Tweets were streamed in real time with the API of Twitter and a string was created to filter and select the tweets according to the parameters summarized in 1. Each Leather word is combined with CSR words and the same with Hashtags. The result is a long list of word combinations that will ensure the tweets found are from the appropriate topic.

2.3. Algorithm

Twitter streaming process incorporates two important set of rules to be followed. In the first place the implementation of the algorithm *streaming_connection* is required to perform tweet streaming 24 h a day, every day of the week to make sure we do not lose any tweets according to the rules established. Secondly, the algorithm called *updating_connection* is needed to update metrics. The proposed method not only considers that both algorithms run in parallel, but also accounts to exchange information through a database between them.

The function called *create_header()* must be placed in the header of all queries in order to get access to TWITTER'S API through a token provided by this platform.

Algorithm 1 Streaming connection algorithm

while true **do**
  $headers \leftarrow create\_headers()$
  $rules \leftarrow get\_rules(headers)$
  $delete \leftarrow delete\_all\_rules(headers, rules)$
  $new\_rules \leftarrow set\_rules(headers, delete)$
  $get\_stream(headers, new\_rules)$
end while

Algorithm 1 deletes the old rules, inserts new ones, listens for incoming data streams, and stores them into the database. The responsible for deleting and inserting new rules to the API session are *get_rules*, *delete_all_rules* and, *set_rules*. First and foremost challenge was to get enough data quality. Several times the rules had to be readjusted to ensure inaccurate data that has nothing to do with the leather industry do not access to the database. This type of data ranges from explicit content to video-game items. The function *get_stream* is process-blocking, and it only unblocks if the connection is lost. Once the process is unlocked, it iterates again to recover the connection failure.

In other words, the algorithm 1 identifies those tweets that are going to be analyzed with further detail without extracting the important information such as metrics, type of publication, etc. For this reason it is required a second algorithm that can obtain this key data for a further sentiment analysis.

Algorithm 2 Updating connection algorithm

```
while true do
    headers ← create_headers()
    ids ← gather_ids()
    for all id ∈ ids do
        update(headers, id)
    end for
end while
```

As noted above, algorithm 2 was needed to fill the gaps. This algorithm retrieves tweets from the database and updates them.

Beforehand, the IDs[6] were created through the Twitter's API, and they were stored by algorithm 1. The function *gathering_ids* queries every single document generated by algorithm 1 and list each ID.

After that, the algorithm 2 queries the missing information to the API and saves the updated information in the same database. This algorithm also compiles information regarding the user or content creator taking into account if is a tweet, retweet or a reply. In these last two cases it also saves the same information

[6]Each object within Twitter - a Tweet, Direct Message, User, List, and so on - has a unique ID.

from the original or source tweet in a different database collection.

2.4. Analyzer

After the processing of all the data collected and the stemming of the text we present the method used to solve the problem of the classification of texts by their feeling at the document level and graphics generation to summarize the results in a readable way.

The analyzer makes use of an API powered by Azure (Microsoft) called Text Analytics, developed to analyze a text and return the overall sentiment. The possible outcomes from this API are labeled with the category of feeling to which they belong such as positive, negative, neutral, or in case of uncertainty, mixed.

We considered that there were important benefits from using this Artificial Intelligence (AI), namely, it does not require machine-learning expertise, it is heavily trained by Natural Language experts, including it is trained outside the samples to reduce the likelihood of bias.

The analyzer consists of one main process in charge of sending data coming from the hub to the Text Analytics API, receive the calculation result and transform them into graphs.

This is schematically shown in figure 3, the results obtained graphically are stored inside the container as raw files due to all results are graphics. As a result, a database solution was discarded due to, in principle, can give rise to memory management constraints, and a data center was also dismissed because of its complexity.

The Analyzer process may not be fast because it depends on how many preprocessed data is available. The larger it is, the more time it will take, bearing in mind it has to be tolerant to internet disconnections since it is using an external API for the sentiment analysis calculations.

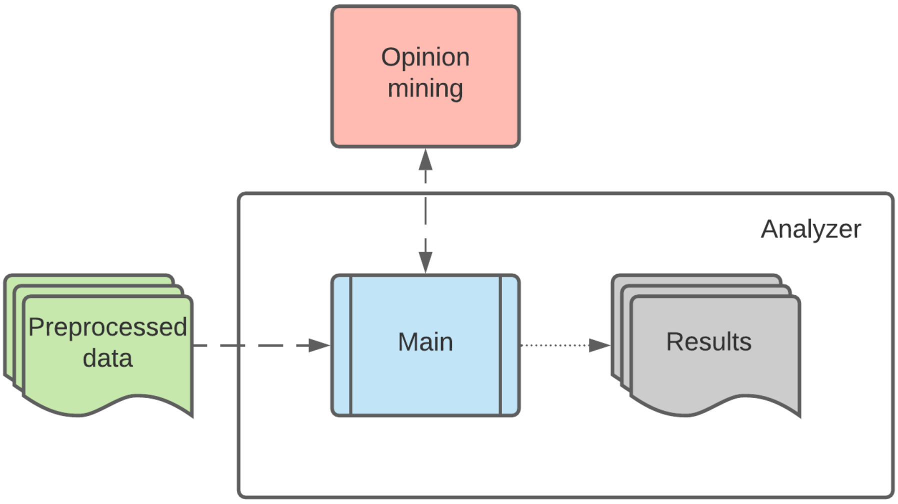


Figure 3: Analyzer's structure

Algorithm 3 gives us a brief idea of the global structure we have applied to get the tweets' data from the database. This is executed by the function called *get_tweets_idtext* which returns a list of dictionaries[7].

Algorithm 3 Analyzer algorithm

$textids \leftarrow get\_tweets\_idtext(database)$

$texts \leftarrow get\_tweets\_text(textids)$

$generate\_wordcount\_csv(texts)$

$build\_wordcloud()$

$calculate\_sentiment\_analysis(textids, database)$

$pie.generate(database, log)$

These dictionaries only have the id and the text as keys. As we were

[7]Dictionaries are Python's implementation of a data structure that is more generally known as an associative array of key-value pairs. Each key-value pair maps the key to its associated value.

only interested with the texts for illustrations purpose we used the function *get_tweets_text* made from *textids*. After counting all words the function *calculate_sentiment_analysis* takes action calculating the opinion of each ID in order that the *pie.generate* can create the graphics for the according sentiment analysis. Technologies used for plotting the graphics are mathplotlib for purely standard graphics and wordcloud for fancy word clouds.

Algorithm 4 Calculate sentiment analysis function

$full\_tweets \leftarrow filter\_tweets(textids)$

$filter\_by\_asso(full\_tweets)$

$full\_ids \leftarrow withdraw\_ids(full\_tweets)$

$full\_metrics \leftarrow aggregate\_metrics(full\_ids, database)$

$calculate(full\_metrics, database, log)$

Algorithm 4 filters from the whole database those tweets which at least contain one set of words that belongs to the row *word* in the table 1.

2.5. Framework overview

Despite the jargon, the framework is simple. Figure 4 shows the global structure and the toolkit used for the streaming of tweets and the further analysis. We considered it was needed to set an environment that isolates the program from the hardware. One of the simplest ways to do it is running all elements of the program with containers [8] where the database is in the middle between components.

3. Results

This section summarizes and discusses the main findings of the work.

[8] A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another.

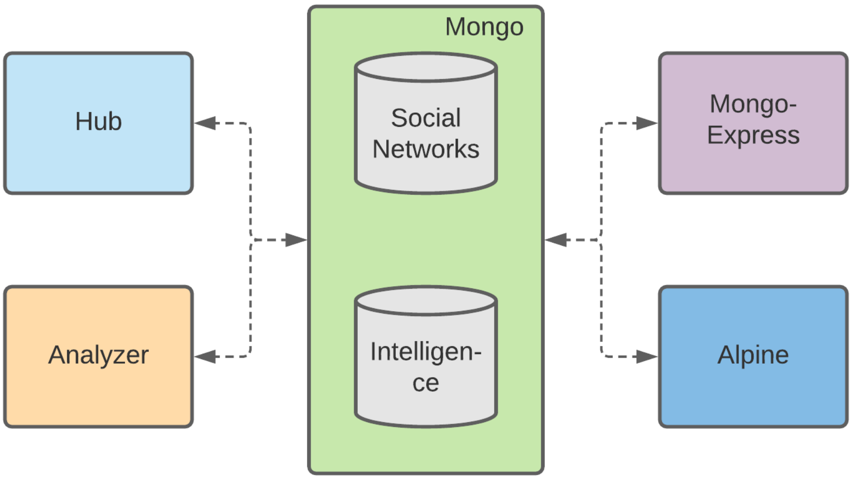


Figure 4: Micro-service framework overview

3.1. Word Clouds

The roots of the words were used to get a more accurate result of the filtered words. For example, the root "leather" refers to all the words that are written by those letters, as its plural "leathers". These words or combination of words (i.e. "leathers", "sustainable leather", "fashion leather") are the ones that Twitter users use when they talk about CSR in the leather industry, which indicates that when they hear about these terms those are the first ideas with which they relate the CSR. Figure 5 gives a general vision of which words do people use when talking about topics related to CSR in the leather industry. The most used word after leather (Because the research scope is on the leather industry) is *fashion* since the sales of leather goods and shoes have expanded greatly during the past decade given by most companies from these industries have turned leather into a seasonal fashion Colucci et al. (2020). It is followed by *Environment* and *sustainable* which is clearly illustrated that the connection between them and leather is less strong however if we extrapolate results we can conclude that when people think about leather the second topic that comes to their mind is which impact have against the planet and the economy. These results are consistent with different studies published in the literature that reveal the extensive interest on the topic of curbing environmental pollution of the manufacturing process in the leather industry China et al. (2020) Morera et al. (2007) de Aquim et al. (2019), materials life cycle analysis Joseph and Nithya (2009) and green chemistry approaches Sathish et al. (2016). Whatever the reasons, and despite the huge improvements made by so many tanners around the world and the academic research, the leather industry has not moved ahead of this preoccupation Redwood (2013).

Figure 5: Word-clouds without any correction

We thought that study data might be supplemented by removing the word leather and thus detect weaker links. This is depicted in Figure 6 which demonstrates that doesn't exists too much connection between *leather* and *tanneries* as well as a weak link between *leather* and finished products.

The results obtained in the last figure 7 may have implications for understanding where the CSR of the leather industry is shifting. More and more people is demanding greener alternatives to the traditional leathers. These results are also in line with previous researches. For instance, Choi and Lee (2021) studies consumers' posts in Google and Twitter focusing on the transition of

Figure 6: Word-clouds without the leather word

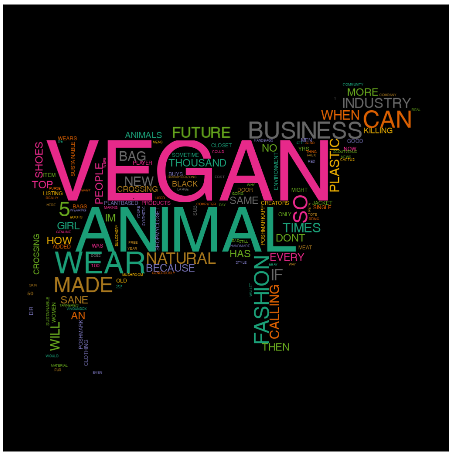


Figure 7: Word-cloud without filters

global consumers' awareness toward vegan materials. Chong Liao and Solatorio (2019) compares a synthesized bio-material in the lab with animal and imitation leather. Developing biotextiles has become fashionable, with novel biodegradable materials developed from waste pineapple leaf fibres (Piñatex™), fungi or even cactus garnering "likes" and shares on social media. It is worth mention that it is deceiving to call a material 'Vegan Leather', but this is often done to market the material to imply it has the beauty, natural quality and long lasting benefits of leather. 'Vegan leather' is usually either artificial or synthetic, or one of a new variety of alternative materialsNaturally (2022).

3.2. Sentiment analysis of tweets

The results obtained with the user and the TEXT ANALYTICS API show in Figure 8 there is a clear tendency to classify the tweets as negative (around 43%). Approximately one of every five (20%) tweets is considered mixed, while positive tweets account approximately just for 8%.

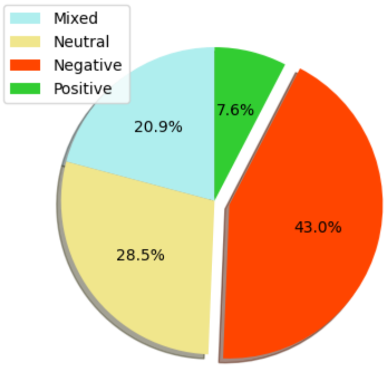


Figure 8: Sentiment classification of the total Tweets.

Figure9 shows the results by CSR topic. Due to the wide variety of words, it has been chosen a set of seven CSR key words well distanced in meaning so there is no overlapping between them. In the interest of clarity, it may be mentioned that the sentiments are classified in four categories; positive, negative, neutral and mixed. One sentiment is counted as mixed when falls into two different categories. As an example, a positive sentence plus a neutral one in a tweet is counted as mixed.

In addition there is a concluding result that indicates in which topics (economy, human rights, sustainability...) CSR fails and where it succeeds. *sustain-*

abilty and *transparency* are less favorable in the public eye whereas *green*[9] and *human rights* are more favorable.

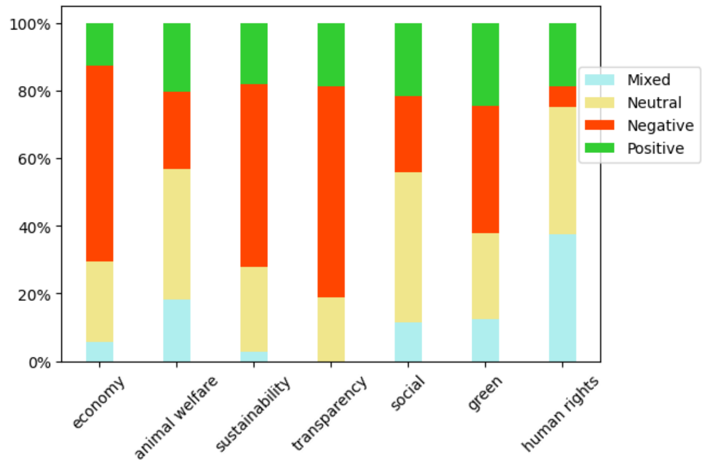


Figure 9: Opinion related to topics

According to the figure 9, there is evidence to support the hypothesis that:

"The general opinion is less favourable, lean towards neutral."

All columns appear to confirm the hypothesis, however the SN might contain bots[10] and most of these negative tweets belong to people with negligible influence. Consequently the results must take into account the engagement perceived. Otherwise, it would not give a clear picture of the opinions because it gives equal weight to all tweets. And as a consequence, it would not confirm the hypothesis.

Retweets and likes are indicators of the impact of tweets on society. In Figure 10 we take into account retweets, likes, replies and quotes when weighing the results. For this reason, figure 10 is clearly different compared with figure 9. It

[9]Word related to friendliness with the environment.

[10]Definition: Software that controls a Twitter account via the Twitter API. It may autonomously perform actions such as tweeting, re-tweeting, liking, following, unfollowing, or direct messaging other accounts.

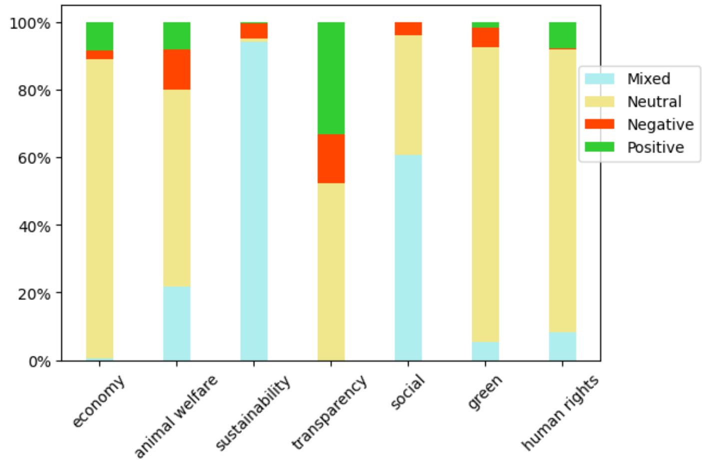


Figure 10: Opinion related to topics weighted

can be assumed that figure 10 has a consensus of neutrality with the exception of *sustainability*. On the other hand, it is important to underline that the positive tweets outweighs the negative ones except the word *social* and *sustainability*. With the weighted result in hand, it is imperative to re-frame the hypothesis so it can be supported by figures 9 and 10. The new hypothesis asserts:

> "By giving the same significance to all opinions, the leather industry is thought mainly negative. Notwithstanding, the opinions that generate more impact in the public eye are prone to neutral."

3.3. Users classification

Figure 11 depicts the percentage of tweets written by anonymous personal accounts (that is, public in general) and those written by official accounts (that is leather companies, leather associations and CEOs of these companies or associations). The amount of "anonymous" (*not verified*) users is normal in this SN because to become a *verified* user, one has to have a certain amount of followers, or reputation. Furthermore, that explains why negative tweets nearly always eclipse in number the neutral ones. This means, that "anonymous" user tweets criticize on the leather industry.

Results show that only a small amount of tweets (around 2%) are from official accounts. These type of accounts generally have more followers and more influence on society than anonymous personal accounts therefore they should increase the adoption of new strategies to reach stakeholders satisfaction and transparent results reporting. According to a report published by eur the need for the leather sector to use social media and communicate effectively has become increasingly important since it opens up windows of opportunity for businesses to disseminate information and to engage stakeholders Cho et al. (2017) Araujo and Kollat (2018).

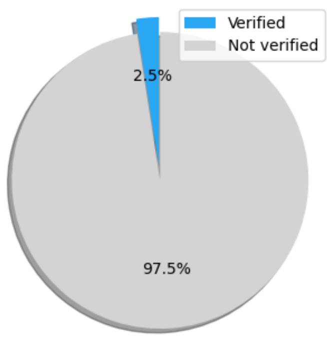


Figure 11: Users classification.

Some general studies on the disclosure of CSR in social media can be found in the literature. For instance, Morsing and Schultz (2006) affirms that social media platforms favor the involvement and dialogical strategies based on CSR activities. Furthermore, social media communication encourages stakeholders to participate in the content provided in those platforms by companies Dellarocas (2003). As a consequence, companies gain more credibility and competitive advantage through the interaction with the stakeholders Bilic (2010) Fieseler et al. (2010). Specific studies referring to the leather sector are very scarce.

4. Discussion

4.1. Implications for theory

In accordance with previous studies published in the literature Śmiechowski and Lament (2017) Omoloso et al. (2021) Abebe (2020) Moreno and Capriotti (2009), we find empirical evidence that CSR in the leather sector is an increasingly studied topic by different researchers and is a core strategy of many companies. While previous studies published in the literature have mainly used traditional methods for data collection such as reports, interviews or questionnaires, this investigation has introduced the use of big data, which opens up new opportunities for research and provides a valuable source of information. To take advantage, CSR researchers and practitioners need to be familiar with computational data collection techniques such as APIs and programming algorithms. Although the project has third party dependencies, the project is easy to maintain and update as it makes use of API calls. By updating those calls, the project should be running again. Additionally, the project has been developed independently of the hardware meaning that hardware compatibility will not be a limitation. Twitter is the perfect data source due to maximum tweet length allowed as too long documents have noise features that hinders the sentiment analysis accuracy Choi and Lee (2017). This research has made a substantial contribution in the engineering part establishing a base where further opinion mining developers can rely on.

The results demonstrated in this work provide a new database of tweets for the leather sector that will serve for future projects in this area helping to understand which are the most spoken topics about leather's CSR words, which are the opinions about them and where the industry is shifting in terms of CSR.

4.2. Implications for practice

Despite having highly valued features and properties, leather is just perceived as a product; a bag, a sofa, a jacket, but it seems to have no value in itself. From the point of view of the end user the leather goes straight from the slaughterhouse to the luxury shop window and everything in between is the absolute darkness. Leather companies face a challenging competitive environment and must adapt and realign strategies to continue responding positively to environmental and sustainable issues. Fashion industry probably has been the 'Savior' of leather during last decades and major brands have had and still have control of all marketing and perception of leather in their hands. This indicates a need to take action in favor of leather industry in order to decide what is good and wrong, which skin or hide counts and provide added value and meaning to the manufactured leather.

As Lattemann and Stieglitz (2007) state, social media connects companies with their stakeholders and engage them to discuss opinions on current topics and trends. In this sense, the results of the study presented in this paper show that some leather companies have started to use social networks to provide a viable source of information and to create an environment for effective stakeholder dialogue in CSR discussions on social network platforms as Twitter.

From the business side, the project has clarified some questions regarding public opinion and the main topics discussed on the leather industry related with CSR. This is a valuable source of information for the CSR strategy definition. A key strength of this research lies within the fact that this framework provides a platform for further research and application to other sectors.

4.3. Limitations of the study and further research

Our research has some practical limitations. The sole focus on the word leather when applying Twitter's filters leaves other industry-related words discarded that may add value to this research. There are identifiable shortcomings with the Twitter API due to the constraints in quantity, length and structure of the rules. Due to time limitations, only one inbound source of information is processed. Nonetheless, the service has a structure that allows the integration of more than one incoming source of information and thus compare results between different social networks. It would be also interesting to conduct similar studies analysing correlations between topics and words and to correlate topics and profiles of users.

5. Conclusions

CSR engagement of firms in controversial industries has been a topic of great interest in recent years for shareholders, regulators, and academic researchers. Leather sector has been included by the society under this sinful umbrella due to they have not made themselves known and its lack of communication to justify or contradict all the negativity focused against them. The results obtained in the investigation are consistent with the fact of not being transparent and communicative generates a multiple negativity in Social networks.

Communicating effectively CSR activities benefits company stakeholder engagement and may turn into positive attitudes from stakeholders towards the company. In this sense, some leather companies have started to use social networks to inform stakeholders and to promote their CSR practices, and now, different stakeholders play an active role in CSR discussions on social networks sites such as Twitter.

Leather sector needs to communicate better their contribution to upcycle an unavoidable waste from the food industry, to produce a versatile, durable, unique material, ideal for the circular economy that the world must move towards. However, these same materials are often dismissed through a lack of understanding of the manufacturing process and its supply chain, or through the application of questionable science generally in the form of incomplete and incomparable or out-dated Life Cycle Assessments, and the marketing of new, often fossil fuel-based materials claiming unsubstantiated levels of sustainability. In this sense, the examination of big amounts of data obtained from social media may help in the aim of analyzing the CSR debate in order to communicate with its stakeholders more effectively.

This project has achieved the objectives proposed. We have developed a big data tool that mines opinions about the leather industry's CSR aimed to understand what topics of CSR in the leather industry are discussed in the

Twitter sphere.

In the course of this work we discovered that fashion is linked in a strongly manner with leather, probably due to industries have turned leather into an all-season product. The debate is related with the impact of leather industries against the planet and the economy and many stakeholders advocate for more environmental responsibility since they consider vegan alternatives such as organic bio-materials that are made of anything except animals. On the other hand, the percentage of tweets posted by anonymous personal users (public in general) accounts for 97,5% while those written by official users only reach 2,5%. These findings indicate that the leather sector should improve the CSR disclosure and adopt a better stakeholder engagement strategy. There are several studies that confirm social networks can lead to positive outcomes since they are not just a way to advertise what company does in regards to CSR, but also to receive input from others for their CSR activities.

Results also reveal a clear tendency to negative sentiment tweets (around 43%) in front of mixed tweets (20%) tweets and positive tweets account approximately just for 8%. *Sustainabilty* and *transparency* are less favorable in the public eye whereas *green* and *human rights* are more supportive.

Acknowledgments

This work was supported by the Ministerio de Ciencia e Innovación under contract PID2020-113614RB-C22 funded by MCIN/AEI/10.13039/501100011033. Also, the authors are members of the research group 2017-SGR363, funded by the Generalitat de Catalunya.

References

, . Social & environmental report 2020. URL: https://www.euroleather.com/doc/SER/European%20Leather%20Industry%20-%20Social%20and%20Environmental%20Report%202020%20-%20EN%20web.pdf.

Abebe, M.G., 2020. Corporate social responsibility and financial performance: Evidence from the ethiopian leather industry. Journal of Commerce and Accounting Research 9, 50.

Alharbi, A.S.M., de Doncker, E., 2019. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cognitive Systems Research 54, 50–61.

de Aquim, P.M., Hansen, É., Gutterres, M., 2019. Water reuse: An alternative to minimize the environmental impact on the leather industry. Journal of environmental management 230, 456–463.

Araujo, T., Kollat, J., 2018. Communicating effectively about csr on twitter: The power of engaging strategies and storytelling elements. Internet Research .

Bilic, I., 2010. Development of web corporate communications function by official web sites and value added ranking: Case of croatia. The Business Review, Cambridge 15, 151–158.

China, C.R., Maguta, M.M., Nyandoro, S.S., Hilonga, A., Kanth, S.V., Njau, K.N., 2020. Alternative tanning technologies and their suitability in curbing environmental pollution from the leather industry: A comprehensive review. Chemosphere 254, 126804.

Cho, M., Furey, L.D., Mohr, T., 2017. Communicating corporate social responsibility on social media: Strategies, stakeholders, and public engagement on corporate facebook. Business and professional communication quarterly 80, 52–69.

Chodorow, K., 2013. MongoDB: the definitive guide: powerful and scalable data storage. " O'Reilly Media, Inc.".

Choi, Y., Lee, H., 2017. Data properties and the performance of sentiment classification for electronic commerce applications. Information Systems Frontiers 19, 993–1012.

Choi, Y.H., Lee, K.H., 2021. Ethical consumers' awareness of vegan materials: Focused on fake fur and fake leather. Sustainability 13, 436.

Chong Liao, C., Solatorio, N., 2019. Synthesis of Cellulose by Acetobacter Xylinum: A Comparison Vegan Leather to Animal and Imitation Leather. Ph.D. thesis. University of Wyoming. Libraries.

Colucci, M., Tuan, A., Visentin, M., 2020. An empirical investigation of the drivers of csr talk and walk in the fashion industry. Journal of Cleaner Production 248, 119200.

Dellarocas, C., 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management science 49, 1407–1424.

Fieseler, C., Fleck, M., Meckel, M., 2010. Corporate social responsibility in the blogosphere. Journal of business ethics 91, 599–614.

Joseph, K., Nithya, N., 2009. Material flows in the life cycle of leather. Journal of Cleaner Production 17, 676–682.

Kanagaraj, J., Senthilvelan, T., Panda, R., Kavitha, S., 2015. Eco-friendly waste management strategies for greener environment towards sustainable development in leather industry: a comprehensive review. Journal of Cleaner Production 89, 1–17.

Kerrisk, M., . crontab(5) - Linux manual page. URL: `https://www.man7.org/linux/man-pages/man5/crontab.5.html`.

Lattemann, C., Stieglitz, S., 2007. Online communities for customer relationship management on financial stock markets-a case study from a project at the berlin stock exchange .

Moreno, A., Capriotti, P., 2009. Communicating csr, citizenship and sustainability on the web. Journal of communication management .

Morera, J.M., Bacardit, A., Ollé, L., Bartolí, E., Borràs, M.D., 2007. Minimization of the environmental impact of chrome tanning: A new process with high chrome exhaustion. Chemosphere 69, 1728–1733.

Morsing, M., Schultz, M., 2006. Corporate social responsibility communication: stakeholder information, response and involvement strategies. Business ethics: A European review 15, 323–338.

Naturally, 2022. Leather & leather alternatives - a guide to labelling. URL: https://www.leathernaturally.org/Education/Fact-Sheets/Guides/Leather-leather-alternatives-a-guide-to-labelling.

Omoloso, O., Mortimer, K., Wise, W.R., Jraisat, L., 2021. Sustainability research in the leather industry: A critical review of progress and opportunities for future research. Journal of Cleaner Production 285, 125441.

Pons, A., Vintrò, C., Rius, J., Vilaplana, J., 2021. Impact of corporate social responsibility in mining industries. Resources Policy 72, 102117.

Redwood, M., 2013. Corporate social responsibility and the carbon footprint of leather. J. Soc. Leather Technol. Chem. 97, 47–55.

Redwood, M., 2018. URL: https://www.leathernaturally.org/getattachment/Latest/Featured-Articles/Education/White-Paper-Explores-Sustainability-of-Leather/White_Paper_Sustainability_Leather_Naturally_Download.pdf.aspx?lang=en-US.

Sathish, M., Madhan, B., Sreeram, K.J., Rao, J.R., Nair, B.U., 2016. Alternative carrier medium for sustainable leather manufacturing–a review and perspective. Journal of cleaner production 112, 49–58.

Śmiechowski, K., Lament, M., 2017. Impact of corporate social responsibility (csr) reporting on pro-ecological actions of tanneries. Journal of Cleaner Production 161, 991–999.

Thelwall, M., Buckley, K., Paltoglou, G., 2011. Sentiment in twitter events. Journal of the American Society for Information Science and Technology 62, 406–418.

Label	Words
ind_words	raw material, leather, tanning, tanneries, tannins, tanning agents, hide, skin, finished leather, wet blue, wet white, pelt, solid waste, water treatments, footwear, clothes, furniture, calfskin, raw materials, chrome, dyeing, fashion, sustainability, bags, luggage, garments, leather craft, leather goods, sustainable fashion, animal welfare
hashtags	handcrafted, leather, future, leather, safety, first, environment, jobs, development, diversity, renewable, safe, respect, traceability, rawhide, leather, goods, awareness, global, goals, sustainable, fashion, leather, naturally, water, management, carbon, neutral, environmental, pollution, recycle, ethical, fashion, zero, waste, clean, tech, vegetal, circular, economy, vegetal, leather
csr_words	CSR, ethic, ethics, ethical, responsible, responsibility, legal compliance, care, respect, reputation, social, socialgood, community, charity, philanthropy, volunteer, nonprofit, culture, sustainable, sustainability, environment, renewable, climate, development, rehabilitation, stakeholder, code conduct, eco, green, safety, safe, diversity, gender parity, transparency, labor practice, human rights, shared value
companies	Horween Tannery, Shinki Hikaku, S.B. Foot Tanning Co., Hermann Oak, J&FJ Baker, DaLuca Straps, Harles F. Stead Leather, Le Farc, Badalassi, Guidi, Tannerie d'Annonay, Henan Prosper, JBS, scottishleathergroup, eccoleather, gruppodani, mastrotto, nquart, krismikita, stahl, quimser, boltthreads, pinatexofficial, desertopelle
agents	Leatherworkinggroup, Eco2l, oeko-tex

Table 1 Main Filters that we used to stream tweets.

Corporate Social Responsibility of the leather tanning sector

Adrià Pons[a], Josep Rius[a,1], Carla Vintró[a], Sergi López[b]

[a]Dept. of Business Administration, University of Lleida, Jaume II 73, 25001 LLeida, Spain.
[b]Dept. of Computer Science, University of Lleida, Av. Jaume II 69, 25001 LLeida, Spain;

Abstract

Reasons to adopt Corporate Social Responsibility reporting by leather tanning companies must primarily be sought in the specific nature of the industry, where leather is a material from renewable sources, however part of its manufacture process might pose one of the gravest environmental hazards in the world. The tanning industry appears, therefore, as a sector compiling CSR reports expected to confirm its environment-friendly actions. The purpose of this research is to examine and understand CSR roles and diffusion in the small and medium enterprises of the leather tanning industry in Catalonia (Spain), especially in relation with the adoption of other management systems. It identifies the most distinguished CSR practices, procedures and metrics and the profiles of companies more prone to adopt them. We used a qualitative approach to fulfill this aim. Results show that leather companies are familiar with CSR practices, but there is further scope for improvement in terms of formalization of CSR procedures and measurement systems. Results show that the majority of socially responsible practices are related to environmental issues and consequently there is a relation between CSR and the application of environmental management systems. The analysis reveals that the management of CSR activities improves with the diffusion of knowledge on CSR practices.

Keywords: CSR; Sustainability; Tanning; Leather industry; Environmental management

[1]CONTACT Josep Rius. Email: josep.riustorrento@udl.cat. Phone: +34 973 702 711

1. Introduction

Leather sector is a world-wide leading market as it provides raw materials for a wide range of consumer goods. The principal destination use of leather has traditionally been the footwear sector, that is still the first client in Europe with a 38% share on the total. Nevertheless in recent years, other important destinations increased, such as leather goods (22%) and car interiors (13%) eur (b). The tanning process plays a fundamental role concerning the processing of skin or hide of animals into a durable material preventing its decay and making it persistent and flexible. Leather making is a very old traditional process that serves social needs, employs many skilled/unskilled persons while being one of the highest contributors of global economy Kanagaraj et al. (2015). Today, with more than two thousand years of history, the tanning industry in Europe represents a strategic segment of the manufacturing sector, thanks to the combination of tradition and continuous innovation. At national level European tanneries present different characteristics, depending on their productive specialization. The most important tanning sectors of southern Europe, such as Italy, Spain, France and Portugal, are mainly composed by small and medium enterprises (SMEs) and they are mostly specialized in the production of leather for the fashion industry. Such leather often needs an artisanal approach that big companies are not always able to provide. Conversely, the tanning sectors of central and northern Europe (Austria, Netherlands, Germany, Sweden, UK) present higher average dimensions of companies as scale economies play a key role for their productions. The average size of an EU tanning company is currently 21 employees and it is important to notice that the rate was 24 in Y2000. This tendency seems to support the slogan 'Small is beautiful', suggesting that those companies have been better in reacting to the big changes that the industry had to face since the beginning of the new millennium. The European leather production has always been characterized by its capability of processing the hides and skins of all the main animal typologies and of serving customers in all the main manufacturing applications. As a matter of fact, about 99% of the

EU leather production constitutes the recycling of animal by-products, residues of the meat industry. Complementarily, the niche of exotic leather concerns a tiny share in terms of square meters (about 1% of the total) but it is quite relevant in terms of value. Such leathers are highly requested by clients of the luxury segment eur (b) eur (a).

On a Spanish level tanning is an industrial sector with a long-standing tradition: beginning amongst craftsmen, it has remained in place for centuries, evolving progressively to its current high technological level. This sector is made up by 98 companies employing around 2.300 workers in 2018. The companies geographical distribution shows a strong concentration in the regions of Catalonia, Valencia and Murcia. 2018 tanning industry output reached 690 million euros led by Catalonia, one of the most industrial regions in Spain, concentrating more than the 50% of total workers and companies from this sector ace, however, the impact of COVID-19 on Europe's Leather Industry, its workers and their families, may trivialise the progress of the sector's social accountability and environmental performance. More detailed data can be found (in Spanish) on the Ministry of Industry website (`https://www.mincotur.gob.es/es-es/IndicadoresyEstadisticas/Presentaciones%20sectoriales/03.%20Cuero%20y%20calzado.pdf`).

The importance of sustainability in the leather industry has continually increased over time focusing its research on the environmental aspect Chen et al. (2014); Zuriaga-Agustí et al. (2015). In this sense, the first decade of the 21st century in particular has seen a renewed debate about tanning and its sustainability. This is owing to advances in technology, legislation, standards, corporate social responsibility (hereinafter CSR) and consumer demands Redwood (2013) while fewer published research about the social and economic aspects existence Gupta and Racherla (2018); Śmiechowski and Lament (2017), even though the interconnectedness of the dimensions, environmental endeavours can have multiplier effects on social and economic sustainability and vice versa Omoloso et al. (2020a).

Following the Brundtland definition of sustainable development WCED (1987), the main challenge for the sector is to demonstrate that, despite the fact that leather production is perceived as a polluting industry and its stigmatization, it contributes to upcycle waste (hides and skin) generated from the meat and dairy industries Dixit et al. (2015), Joseph and Nithya (2009), thus reinforcing the "circular economy industry" nature of leather and that companies mind their social dimensions Omoloso et al. (2020a). It is important to remember that the leather Industry market is strictly connected to the meat market. In fact, this industry is able to convert a by-product of the food industry into a noble material which otherwise should be disposed into waste Beghetto et al. (2013). Leather industry has to deal with these challenges and has to assume responsibilities in local and national development. Tannery processes are characterised by the use of significant quantities of chemicals, energy and water Selvaraj et al. (2019) Zuriaga-Agustí et al. (2015) Kanagaraj et al. (2015). As a result, research to improve the sustainability and ethical management of the industry processes has proliferate during the last two decades Omoloso et al. (2020a). Sustainable development and ethical management have been included in the European Leather Industry Partner's Roadmap 2018-2025 cot (2018), and various national and international initiatives have developed frameworks for sustainability. For instance, 2030 Agenda for Sustainable Development and its 17 Sustainable Development Goals (SDGs) promoted by the United Nations Desa et al. (2016) which look ahead to the opportunities and how to best address the many interconnected challenges of a sustainable recovery at all levels. Another instance is the Confederation of National Associations of Tanners and Dressers of the European Community (COTANCE) has formulated some basic principles of good practice in which they outlined the challenges and opportunities for the sector in four areas: industrial, social/societal, trade and environmental eur (b). The adoption of socially responsible practices – i.e. environmental protection; dialog with stakeholders; community-investment programs; commitment to the rights of employees in the company's policies, etc. Śmiechowski and Lament (2017) Braam et al. (2016) Thijssens et al. (2016) - helps companies to maintain an image of commitment towards the environment and their stakeholders, and to lead the pursuit of a social approval to operate, which is aimed to secure the support of local communities. The social dimension, together with the environmental and economic dimensions, shapes the triple bottom line Kleine and Von Hauff (2009), which is at the basis of corporate social responsibility Elkington (1997).

CSR may be defined as the group of actions that are undertaken by an organization in order to accept the responsibilities resulting from the impact of its activities on society and the environment ISO 26000:2010. This new business approach has special significance in tanning activities. Globalisation has exposed human rights violations and environmental pollution in deprived parts of the world, and the reporting of these unacceptable practices often paints the whole sector with the same brush increasing public concern about CSR over the last years. The writings by the influential journalist Lucy Siegle Siegle (2011) are usual: *'But as well as leather being one of the most prevalent materials in our wardrobe, its production and processing phase is one of the most polluting systems that humankind has managed to come up with'*. She continues with a description of a visit to Uttar Pradesh in India and describes current issues with tannery pollution getting into the Ganges. The number of places in the world that people may travel to see unsatisfactory tanning sights is sadly quite large and does extra harm when built into a narrative by the anti-leather and anti-animal lobby on the internet. These are issues which are approached by pressure groups which use the most striking style to push their opinions out to a narrow audience. According with Redwood (2013) the value of the above-described approach is questionable up until now. The leather industry has chosen to ignore this activity yet there is considerable evidence of spill-over into the general public and perhaps more importantly into key user groups. In order to overcome the potential effects of these kind of negative recognition most companies are engaging in both CSR communication and implementation Rao et al. (2003). The debate surrounding CSR in the leather tanning industry has gained considerable attention with in the academic community since sales of leather goods and shoes have expanded greatly during the past decade due to companies have turned leather into a seasonal fashion Colucci et al. (2020). The discussion about CSR has emphasized its commitment to social responsibility as an emerging topic, and literature seems to indicate that tannery companies have increased their environmental and social consciousness Redwood (2018). In this sense, this 21st century in particular has seen a renewed debate about tanning and its sustainability owing to public concern about the current degradation of the environment and about social actions in the communities affected Thanikaivelan et al. (2005). Different authors have analyzed the topics relating to CSR reporting in the leather sector globally. For instance, Omoloso et al. (2020b) has outlined why sustainability reporting provides several benefits for leather supply chain companies. Laurenti et al. (2017) has concentrated on carbon foot printing of leather, and Śmiechowski and Lament (2017) compared the impacts of CSR reporting on pro-ecological actions between large and small tanneries. Other authors have studied the impact of CSR on local communities Restrepo-Collazos and Castaño-Martínez (2021) Awan et al. (2019) and have analyzed the practices involved in different countries Abebe (2020) Belal (2001) Bimir (2016).

In Western developed countries companies have started to carefully consider the social, ethical and environmental effects of their actions. Many Spanish companies, especially the largest ones, have introduced CSR policies, and the adoption of business ethical practices is increasing. CSR is also promoted by the European Union eur (a). The aim of this paper is to find out whether the leather industry is also committed to CSR by focusing in the leather tanning supply chain industry in Catalonia.

Although the voluntary CSR initiatives are varied, many of them relate to the areas of quality, environmental management and occupational safety Munasinghe and Shearer (1995), areas which may in turn be managed according to international standards such as ISO9001(quality), ISO14001(environment) and OHSAS18001 (health and safety at work), to cite the most common ones. In Europe, and especially in Spain, such standards have been adopted by many companies, from the smallest ones to the largest ones Casadesus et al. (2008).

According to Vintró et al. (2012) the success of management systems makes us think of them as a means for the introduction of CSR practices, even more when CSR is also supported by international standards (ISO26000:2010 Guidance for Social Responsibility).

In the next section, the paper explores the current discussions on CSR in the leather industry in order to gain insight into the positive and negative contributions of leather tanning to sustainable development. The research methodology is then introduced. Next, the results obtained are presented. Discussions and conclusions follow.

2. CSR in the leather industry

Before society had access to glass, paper, rubber, plastics and many modern textiles leather was the most essential material for industry, warfare and many aspects of everyday life. It has constituted one of the world's largest industries for most of man's time on earth Redwood (2013). Long before modern science understood the processes fully the industry used chemistry, biotechnology and technology copied from other areas or from trial and error Covington (2009). Today the manufacture of leather utilizes between 20 and 40 process steps that vary according to the end use of the leather and the preferences of the tanner Sundar et al. (2013), however according to Morera et al. (2007) in all tanning technology operations wastes are generated causing that leather production is defined as an annoying industry and gained broad recognition which lasts to this day as a source of pollution in terms of air, water and solid wastes. Although this might be true in the past, during the last decades several drivers such as advances in technology, legislation, standards, CSR, and consumer demands, have created the right conditions to accelerate a transformation Redwood (2013).

The debate surrounding leather activities and CSR is now centered on sustainability at a global scale Omoloso et al. (2020a). While some authors highlight the positive impacts that leather activities can have Redwood (2013) De Klerk et al. (2019) there is still some kind of nuisance with the link "sustainable development and leather". A major argument against this sector contributing to sustainable development is the nervousness about the CO2 of cattle plus comments about the chemicals and waste involved –i.e. the sportswear company Puma Clark (2012). Moreover, due to the animal origin with a strong link to mistreatment, it is a controversial material with a hostile and harmful social opinion around it. The industry is associated with images of enclosed animals, in poor condition and which are bred and or killed exclusively for their skin.

On the other hand, primary examples of positive contributions of leather to sustainable development and social management include caring for the country, people and culture. These incorporate providing employment to residents of local communities in times when robots and artificial intelligence are eliminating many jobs. Over the last few decades since 1960, leather using industries have helped pull millions of people out of poverty in Korea, Taiwan, India, China and many other countries Redwood (2018). Jakov Buljan of UNIDO discussed this multiplier at an industry conference in 2014 indicating 1 sq. ft. of leather (routinely produced) creates 50 jobs in leather using industries.

As stated by Moktadir et al. (2020), the leather industry can contribute to sustainable development by minimizing the environmental and social impacts throughout its lifecycle, and this is best accomplished through effective CSR management. The author give some guidance on how to improve the sustainability in tanning operations.

Leather activities might be viewed as more socially and environmentally responsible if CSR practices were adopted. In this sense, Śmiechowski and Lament (2017) states that tanneries would need to be subject to a formalized system of environmental monitoring in order to ensure that CSR reporting have direct impact on pro-ecological activities of tanning enterprises. In the literature, there are several studies regarding the extensive interest on the topic of curbing environmental pollution of the manufacturing process in the leather industry China et al. (2020) Morera et al. (2007) Kanagaraj et al. (2015) Dwivedi et al. (2019) Sundar et al. (2002), materials life cycle analysis Joseph and Nithya (2009) and green chemistry approaches Sathish et al. (2016) Krishnamoorthy et al. (2012).

Sustainable development is a priority that has become a common denominator of the corporate policies of European tanneries. It is an irreversible trend, involving all the parties in the supply chain. This is demonstrated by the huge investments that the industry has made over the years and the significant costs that companies have to bear in relation to CSR. This commitment costs on average, 4% of turnover, a figure that grew markedly in the first decade of the 2000s and which has now become a constant on the balance sheets of the sector eur (b).

Closs et al. (2011) defines the Economic sustainability as the "profit" dimension which takes into consideration an organisation's effort to improve the value it generates and delivers to its customers, at the same time reducing the cost of its supply chain related activities. In the context of the leather industry, studies relating to economic sustainability have usually been found in topics relating to customer perceptions and sustainable behaviours when making buying choices De Klerk et al. (2019) Dekhili et al. (2019). Genuine leather products provide several advantages such as quality, good craftsmanship and aesthetic appeal. It is expensive and in many cases much more expensive than products made of man-made textiles Gorp et al. (2012). It is not only bought for its functional value (quality, durability, etc.), but also for social, personal and financial value Kapferer and Michaut (2015) Shukla et al. (2015).

3. Methodology

The basic information for this work comes from a more extensive study that analyses the application of quality, environmental, occupational health and safety and CSR management systems in the leather industry in Catalonia. The research was started with the design of a questionnaire, which was the main source for data collection. A pilot-test was conducted with in a small number of Catalan leather companies guided by the leather cluster of Barcelona in order to prove viability and to detect difficulties in the interpretation of questions. This cluster is a private and non-profit association that represents the interests of the companies and agents of the leather industry in Catalonia. The final version of the questionnaire comprised 25 items concerning the diffusion of CSR practices, CSR procedures, CSR measurement systems and CSR management systems, and the perception of companies about the benefits experimented in their business results. Some of the items were open-ended questions, others were multiple choice questions and some used a five point Likert scale.

Table 1
Technical record of the study.
Source: Questionnaire survey.

Population	48	
Geographical area:	Catalonia (Spain)	
Survey addressed to:	Management Systems Manager (or Business Manager)	
Sample unit:	Company	
Sample size:	48	
Response rate:	39,6% (19 companies)	
Details		Total
Companies included in the leather cluster of Catalonia (year 2020)		48 (100%)
Questionnaires received after the pilot test		3 (6,2%)
Questionnaires received after the first mailing		12 (14,6%)
Questionnaires received after the second mailing		4 (21%)
Total number of questionnaires received		19 (39,6%)

The target population was identified by the "leather cluster of Barcelona (year 20020)". Although a single company may operate in more than one location, the unit of analysis for this research was the company. The population considered was 48 companies. All questionnaires were sent by electronic mail (web-based survey tool) and were addressed to the Management Systems Manager or to the Business Manager. A cover letter explaining the purpose of the study accompanied the survey questionnaire. To increase the response rate, companies were previously contacted by telephone. After the whole process, the deficiencies detected in the answers were rectified through telephone calls or electronic mail.

A total of 19 completed responses were received (39.6% response rate), which reached the targeted overall response rate for a valid assessment since Martínez (1990) observes that response rates of 20% are common and considered satisfactory for post surveys in Spain. Malhotra and Grover (1998) suggest a response rate of over 20% for a positive assessment of mail survey results. The statistical analysis of the data was performed using Google forms and SPSS software. Table 1 shows the technical record of the study.

4. Results and discussion

4.1. Companies' profile

The distribution of the sampled companies is composed by 100% medium or small sized firms according to the European Union Recommendation 2003/361/EC. The quantity of leather produced varies from 20.000 to 3 million square meters a year. These companies mainly concentrate their commercial activity on a international scale.

The literature suggests that bigger companies are more likely to be under public scrutiny Liu and Anbumozhi (2009) and are generally skilled of having better resources Spence (1999). Nonetheless, the study conducted evidence that medium and small companies have also undertaken various socially and environmentally responsible actions (we refer the reader to subsequent sections).

Table 2
Management systems implemented.
Source: Questionnaire survey.

	Management systems			
	Quality ISO 9001	Environment (LWG & ISO 14001)	Health & Safety (OHSAS 18001)	Regulatory compliance
Companies (19)	5 (26%)	7 (36%)	4 (21%)	2 (10%)

4.2. Management systems

Management systems are a set of organizational procedures, responsibilities and processes that help to implement corporate policies regarding quality, environment, health and security, etc. Most usual management systems are in accordance with international certifiable standards such as ISO 9001 (product and service quality management system), ISO 14001 (environmental management system) or OHSAS 18001 (occupational health and safety management system).

About 60% of the surveyed companies described efforts like ISO 9001 certification, ISO 14001 certification or OHSAS 18001 certification to align business practices with the needs and expectations of society. The percentages of application of management systems are detailed in table 2.

Results show that companies have experience in management systems. The number of certifications has increased throughout time, especially when referring to quality and environmental management systems. Customers'competitive pressure Brown et al. (1998) in the first case, and growing environmental regulatory pressure WCED (1987) in the second case, have contributed to this fact.

6 out of the 19 surveyed companies had an specific CSR management system certificate. 3 of them implemented Leather Working Group (LWG) procedures, which measures environmental performance whereas the ISO standard does not. The ISO standard relates to the management system and is explicit in not attempting to quantify environmental performance. The LWG Environmental Audit Protocol is developed by not-for-profit organisation responsible for the world's leading environmental certification for the leather manufacturing industry and assesses leather manufacturing facility operations in absolute terms, for example energy requirement to produce a square foot of leather, proportion of wastes recycled, etc. The others have chosen other options, for instance ZDHC MRSL (Manufacturing Restricted Substance List) focused on measuring indicators such as, wastewater, sludge and air quality validating the work that is being done with chemical inputs and processes. It helps to determine if the out-

put water and air is safer. Another example is the introduction of the United Nations Global Compact initiatives based on CEO commitments to implement universal sustainability principles and to take steps to support UN goals in the areas of Human Rights, Labor, Environment and Anti-Corruption.

The other 40% admitted the existence of certain difficulties in the implementation of management systems. The main considerations had been; time constraints and lack of financial resources.

In spite of these barriers, companies showed an emergent interest in management systems. Most companies (36% of companies that had implemented the ISO 14001 or LWG, 26% for ISO 9001 standard and 21% for OHSAS 18001) admitted that the implementation of management systems had conveyed some benefits. (i.e. better operational implementation and document management). For instance, a surveyed Manager asserted in a questionnaire that:

'Industry environmental regulations have "forced" us to reinvent ourselves, we have to become environmentally aware, the new generations have transformed a business that has been professionalized and modernized rapidly'

4.3. CSR practices

Around 74% of the leather companies participating in the study stated that they knew or had heard about CSR management systems (Table 3). This result led us to high expectations regarding the implementation of CSR practices. In this sense, the actions promoted by public administrations and professional associations may play a prominent role in the diffusion of CSR. Similar conclusions were published in previous studies. For instance, Nijkamp et al. (1999) already considered that public administrations should take a leading role in the promotion of environmental actions. We examined the dissemination of CSR practices. Respondents were required to indicate which of the eight practices detailed in the questionnaire their companies had implemented (See Table 4). Around 60% of the companies had implemented at least three different CSR practices and 16% had applied six or more CSR practices. This is a favorable result and it may even be enhanced in the future. The results in Table 4 mean that companies have committed themselves to ethical and sustainable issues. For example, one of the surveyed companies asserted that:

'Our sector has been improved ostensibly in all its areas: wastewater control, animal welfare, reduction of chemicals, efficient processes, greater energy efficiency, ... But we are afraid that its evolution has not transformed into a change in the perception of the sector'

We combined the all CSR practices into three main groups according to the three vertices of the CSR Munasinghe triangle (environmental, social and economic criteria): CSR management practices linked to environmental management (Fig.1), CSR management practices linked to human resources management (Fig. 2) and CSR management practices linked to sustainable administration of funds.

Table 3

CSR management systems diffusion.

Source: Questionnaire survey.

Do you know...?	CSR management systems
Yes	14 (74%)
No	5 (26%)
Total	19 (100%)

Table 4

Distribution of CSR practices.

Source: Questionnaire survey.

CSR management practices	Companies	%
Codes of conduct	16	84%
Information transparency	5	26%
Promotion of local communities	5	26%
Improved waste management	18	95%
Energy sources consumption control	17	89%
Professional career	4	21%
Cooperation with NGOs	2	10%
Integration of BAT[1] to tanning processes	7	37%
Attendance at seminars and conferences about CSR	6	31%
Reducing consumption of excessive chemical products	1	5%
Total Companies	19	100%

[1]Best available techniques provided by the Spanish Ministry of the Environment in 2013

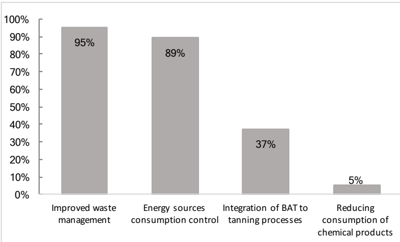


Figure 1: CSR management practices linked to environmental management.
Source: Questionnaire survey.

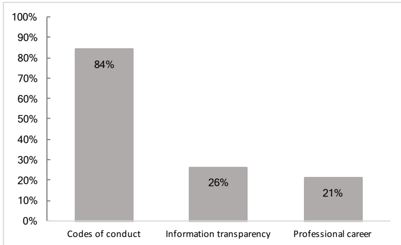


Figure 2: CSR management practices linked to human resources management.
Source: Questionnaire survey.

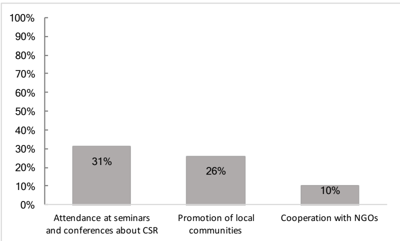


Figure 3: CSR management practices linked to sustainable administration of funds.
Source: Questionnaire survey.

Global results (Figs. 1–3) imply that CSR environmental practices (such as waste management, control of energy consumption and integration of best technology) are more widespread than social actions (i.e. cooperation with non-governmental organizations, promotion of local communities). While this difference could be attributed to a number of reasons, a key element may be the high impact of the leather industry on the environment.

4.4. CSR Procedures and measurements systems

A procedure is a specified way to carry out an activity or process. Procedures describe the aim of the process and its necessary steps, technical elements, technical conditions and human resources. Procedures are included within the documentation of quality, environmental and occupational health and safety management systems. In the same way, procedures are an essential element of CSR management systems.

We examined the application of CSR procedures and measurement systems in the leather companies. Similar work has been considered by Vintró et al. (2012) examining the application of CSR procedures and measurement systems in the mining companies.

Variables 'CSR procedures' and 'CSR measurement systems' were measured with five point (1–5) Likert scales in order to evaluate their application. Descriptive statistics for these variables are shown in Table 5. The mean value (3.052 for CSR procedures and 2.578 for CSR measurement systems), are close to the central point (3) of the scale at the 0.05 level, however suggest that measurement of CSR practices should be improved as it is slightly below. The results obtained in the questionnaire survey indicate a relevant application of CSR procedures (Table 6) however only 10% of the respondents had written procedures for all the CSR activities implemented.

Table 5

Statistical summary of the variables 'CSR procedures' and 'CSR measurement systems'..
Source: Questionnaire survey.

Variable name	Obs.	Mean	Std. dev	Skewness	Kurtosis
CSR procedures	19	3.052	1.223	-0.313	-0,508
CSR measurement systems	19	2.578	1.346	0.423	-0.911

The percentage of CSR measurement systems implementation is relatively low (Table 7) in comparison with CSR procedures. However around half of the surveyed companies (47,3%) use measurement systems regularly.

It is important to note that, in literature, there are some proposals of CSR indicators. For example, Azapagic and Perdan (2000) proposes a general framework with a relatively simple, yet comprehensive set of indicators for identification of more sustainable practices for industry. Wilburn and Wilburn (2013) analyzed a set of sustainable development indicators from several companies

Table 6

CSR procedures diffusion.

Source: Questionnaire survey.

Has the company implemented written CSR procedures?	No. of companies
1: For none of the practices	3 (15,8%)
2: For 25% of the practices	2 (10,5%)
3: For 50% of the practices	7 (36,8%)
4: For 75% of the practices	5 (26,3%)
5: For all the practices	2 (10,5%)
Total companies	19 (100%)

Significant at the 0.01 level

Table 7

CSR measurement systems diffusion.

Source: Questionnaire survey.

Does the company use any metrics in CSR management?	No. of companies
1: For none of the practices	5 (26,3%)
2: For 25% of the practices	5 (26,3%)
3: For 50% of the practices	4 (21,1%)
4: For 75% of the practices	3 (15,8%)
5: For all the practices	2 (10,5%)
Total companies	19 (100%)

based on a global reporting initiative (GRI). Vintro and Comajuncosa (2010) presented another performance chart intended as an internal measure of continuous improvement in CSR. There is still a great deal of work to be done in this area for the leather sector. Hence, future research to investigate sector-specific indicators may be necessary to set the foundation for unification on related practices.

4.5. Perceived CSR utility

Respondents were required to assert whether CSR practices were contributing to improve business results or whether they were not (Fig.4). Variable 'perceived CSR utility' was measured with an open question. Around 78% of the companies perceived a positive impact lead by 'Reduction of environmental management cost' followed by the chance to access to more competitive markets. But not only this, respondents also were required to answer if the implementation of environmental practices contributes to the possibility of reducing the risk of sanctions and litigation measured with a five point (1–5) Likert scale. Almost 90% rated very highly (4-5) this statement.

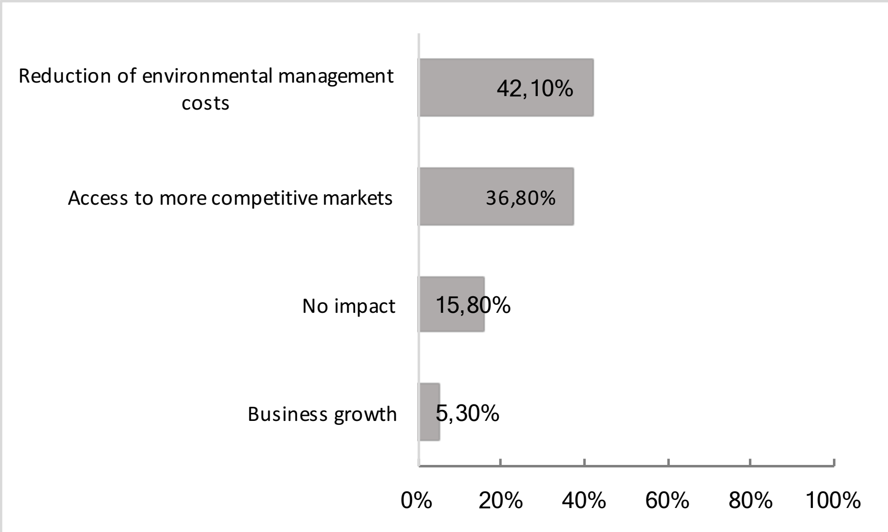
Cross tabulation tables were used in this study to explore the relationship between different pairs of variables: number of CSR practices vs. CSR procedures; CSR procedures vs.CSR measurement system. Spearman rank correlation coefficient and measures of concordance were used in the analysis. The results obtained are summarized in Table 8.

The variable 'number of CSR practices' is significantly and positively associated with the variable 'CSR procedures' at $p < 0.01$. Another variable that shows a positive correlation ($p - value < 0.01$) with 'CSR procedures' is 'CSR measurement systems'.

5. Conclusions

Leather manufacturing sector upcycles an unavoidable waste from the food industry, to produce a versatile, durable, unique material, ideal for the circular economy that the world must move towards. However, it is often dismissed through a lack of understanding of the manufacturing process and its supply chain, or through the application of questionable science generally in the form of incomplete and incomparable or out-dated Life Cycle Assessments, and the marketing of new, often fossil fuel-based materials claiming unsubstantiated levels of sustainability.

Figure 4: Perceived CSR utility
Source: Questionnaire survey.



In order to set lights of consciousness into this darkness, CSR communication has increased over the years, and many companies of different sectors communicate about economic, environmental and social issues in some form, and most of them engage in sustainability reporting. Environmental and social responsibilities are of strategic magnitude for leather activities. As Coppola et al. (2014) states, it is important that companies have a clear understanding of the issues affecting sustainability in the local communities. During the last two decades, investment and finance communities have started to take an interest in climate change, modern day slave labour, corruption and other issues which have come to the force as being major risks not only for the future of business but of society itself and this has forced companies to redefine their strategy. As a result, sustainable development and ethical management have been included in the corporate policies. This paper has examined the diffusion of CSR practices, procedures and measurement systems within the leather industry in Catalonia. Catalan companies from leather industry are small or middle sized, like most

Table 8
Cross tabulation analysis.
Source: Questionnaire survey.

[a]Results	
Pair of variables: number of CSR practices–CSR procedures	
Spearman rank correlation:	0,715
Measures of concordance (p-value: 0.000)	
Kendall's tau-b	0,618
Kendall's tau-c	0,589
Somers' d (number of CSR practices)	0,630
Somers' d (CSR procedures)	0,607
Gamma	0,739
Pair of variables: CSR procedures–CSR measurement systems	
Spearman rank correlation:	0,778
Measures of concordance (p-value: 0.000)	
Kendall's tau-b	0,703
Kendall's tau-c	0,672
Somers' d (CSR procedures)	0,688
Somers' d (CSR measurement systems)	0,719
Gamma	0,829

[a]Significant at the 0.01 level

companies in Europe. Nevertheless does not stop them from implementing management systems. Most companies realize the benefits of having a management system while others admitted unawareness or difficulties to implement them.

Leather companies have heard of CSR (around 74% of the companies were familiar with CSR management systems) and many of them have undertaken some actions, especially in relation with the environment. However, the number of companies that adopt CSR procedures and measurement systems is relatively low, and the diffusion of sustainable leather standards is fairly sparse.

Only 10% of the companies have either written CSR procedures or measurement systems, which constitutes a very low rate of implementation while around 50% of the surveyed companies did not have measurement systems to track the CSR results taking into account that in Spain there is national guide for the Best Available Techniques (BAT) specially created for leather activities (about 20% admitted that they did not know about it).

Findings highlight commitment to CSR: leather companies engage with stakeholders and have implemented different CSR practices oriented to promote the organization's environmental and ethical behavior. According to Cai et al. (2012) this theoretically should provide further benefits to this sector due to CSR engagement is positively associated with firm value in controversial industries.

Leather companies in the sample understand the negative impact of their activities on the environment and they want to observe the law and collaborate with both internal and external stakeholders. For these reasons they have implemented several CSR practices. It is observed that the implementation degree of environmental practices in internal processes is higher than social actions. This result can be attributed to the impacts generated by the tanning processes on the environment which constitute the most perceived threats of leather activities. Cooperation with non-governmental organizations was the least frequent activity. This is also pretty similar to the findings of Vintró et al. (2014) which examines the adoption of environmental practices in the surface mining industry in Catalonia.

Results obtained in the study show that companies with a higher number of CSR practices had implemented more CSR procedures. Moreover, companies with the highest percentages of CSR procedures implementation also described the highest levels of CSR measurement systems application. Consequently, we expect that if the diffusion of CSR practices increases, the formalization and systematization of CSR management will be enhanced.

The results also help to justify that several CSR practices are possible in leather activities, even in small companies. Furthermore CSR may improve the image of leather and tanning companies and help them improve the way they perform their operations. Since companies are used to management systems, this might contribute to the diffusion of CSR management systems and hence the formalization of CSR practices.

Acknowledgments

This work was supported by the Ministerio de Ciencia e Innovación under contract PID2020-113614RB-C22 funded by MCIN/AEI/10.13039/501100011033. Also, the authors are members of the research group 2017-SGR363, funded by the Generalitat de Catalunya.

References

, a. The leather industry in the eu. URL: `https://ec.europa.eu/growth/sectors/fashion/leather/eu-industry_en`.

, b. Social & environmental report 2020. URL: `https://www.euroleather.com/doc/SER/European%20Leather%20Industry%20-%20Social%20and%20Environmental%20Report%202020%20-%20EN%20web.pdf`.

, . The spanish tanning industry. URL: `https://www.leather-spain.com/en/the-sector/`.

, 2018. European leather industry partner's roadmap 2018-2025. URL: `https://www.euroleather.com/doc/TanningRoadmapEN-2018.pdf`.

Abebe, M.G., 2020. Corporate social responsibility and financial performance: Evidence from the ethiopian leather industry. J. Commer. Account. Res 9.

Awan, U., Khattak, A., Kraslawski, A., 2019. Corporate social responsibility (csr) priorities in the small and medium enterprises (smes) of the industrial sector of sialkot, pakistan, in: Corporate social responsibility in the manufacturing and services sectors. Springer, pp. 267–278.

Azapagic, A., Perdan, S., 2000. Indicators of sustainable development for industry: A general framework. Process Safety and Environmental Protection 78, 243–261. URL: https://www.sciencedirect.com/science/article/pii/S0957582000708834, doi:https://doi.org/10.1205/095758200530763. sustainable Development.

Beghetto, V., Zancanaro, A., Scrivanti, A., Matteoli, U., Pozza, G., 2013. The leather industry: a chemistry insight part i: an overview of the industrial process. Sciences At Ca'Foscari .

Belal, A.R., 2001. A study of corporate social disclosures in bangladesh. Managerial Auditing Journal .

Bimir, M.N., 2016. Corporate social responsibility learning in the ethiopian leather and footwear industry. International Journal of Scientific & Engineering Research 7.

Braam, G.J., de Weerd, L.U., Hauck, M., Huijbregts, M.A., 2016. Determinants of corporate environmental reporting: The importance of environmental performance and assurance. Journal of cleaner production 129, 724–734.

Brown, A., Van Der Wiele, T., Loughton, K., 1998. Smaller enterprises' experiences with iso 9000. International journal of quality & reliability management .

Cai, Y., Jo, H., Pan, C., 2012. Doing well while doing bad? csr in controversial industry sectors. Journal of Business Ethics 108, 467–480.

Casadesus, M., Marimon, F., Heras, I., 2008. Iso 14001 diffusion after the success of the iso 9001 model. Journal of Cleaner Production 16, 1741–1754.

Chen, K.W., Lin, L.C., Lee, W.S., 2014. Analyzing the carbon footprint of the finished bovine leather: A case study of aniline leather. Energy Procedia 61, 1063–1066.

China, C.R., Maguta, M.M., Nyandoro, S.S., Hilonga, A., Kanth, S.V., Njau, K.N., 2020. Alternative tanning technologies and their suitability in curbing environmental pollution from the leather industry: A comprehensive review. Chemosphere , 126804.

Clark, P., 2012. URL: https://www.ft.com/content/579f4512-bc62-11e1-a836-00144feabdc0.

Closs, D.J., Speier, C., Meacham, N., 2011. Sustainability to support end-to-end value chains: the role of supply chain management. Journal of the Academy of Marketing Science 39, 101–116.

Colucci, M., Tuan, A., Visentin, M., 2020. An empirical investigation of the drivers of csr talk and walk in the fashion industry. Journal of Cleaner Production 248, 119200.

Coppola, G., Bandinelli, R., Ciarapica, F.E., Dotti, S., Gaiardelli, P., Resta, B., Rinaldi, R., 2014. Building sustainable supply chains in the textile, clothing and leather sectors. Proceedings of the Summer School Francesco Turco , 52–57.

Covington, A.D., 2009. Tanning chemistry: the science of leather. Royal Society of Chemistry.

De Klerk, H.M., Kearns, M., Redwood, M., 2019. Controversial fashion, ethical concerns and environmentally significant behaviour: The case of the leather industry. International Journal of Retail & Distribution Management .

Dekhili, S., Achabou, M.A., Alharbi, F., 2019. Could sustainability improve the promotion of luxury products? European Business Review .

Desa, U., et al., 2016. Transforming our world: The 2030 agenda for sustainable development .

Dixit, S., Yadav, A., Dwivedi, P.D., Das, M., 2015. Toxic hazards of leather industry and technologies to combat threat: a review. Journal of Cleaner Production 87, 39–49.

Dwivedi, A., Agrawal, D., Madaan, J., 2019. Sustainable manufacturing evaluation model focusing leather industries in india: a tism approach. Journal of Science and Technology Policy Management .

Elkington, J., 1997. Cannibals with forks. The triple bottom line of 21st century 73.

Gorp, T.V., Hoffmann, J., Coste-Manière, I., 2012. Brand building: Luxury leather goods brands anatomized. Journal of Global Fashion Marketing 3, 127–134.

Gupta, S.K., Racherla, U.S., 2018. Interdependence among dimensions of sustainability: evidence from the indian leather industry. Management of Environmental Quality: An International Journal .

ISO 26000:2010, 2000. Plastics – Determination of fracture toughness – Linear elastic fracture mechanics (LEFM) approach. Standard. International Organization for Standardization. Geneva, CH.

Joseph, K., Nithya, N., 2009. Material flows in the life cycle of leather. Journal of Cleaner Production 17, 676–682.

Kanagaraj, J., Senthilvelan, T., Panda, R., Kavitha, S., 2015. Eco-friendly waste management strategies for greener environment towards sustainable development in leather industry: a comprehensive review. Journal of Cleaner Production 89, 1–17.

Kapferer, J.N., Michaut, A., 2015. Luxury and sustainability: a common future? the match depends on how consumers define luxury. Luxury Research Journal 1, 3–17.

Kleine, A., Von Hauff, M., 2009. Sustainability-driven implementation of corporate social responsibility: Application of the integrative sustainability triangle. Journal of Business Ethics 85, 517–533.

Krishnamoorthy, G., Sadulla, S., Sehgal, P., Mandal, A.B., 2012. Green chemistry approaches to leather tanning process for making chrome-free leather by unnatural amino acids. Journal of hazardous materials 215, 173–182.

Laurenti, R., Redwood, M., Puig, R., Frostell, B., 2017. Measuring the environmental footprint of leather processing technologies. Journal of Industrial Ecology 21, 1180–1187.

Liu, X., Anbumozhi, V., 2009. Determinant factors of corporate environmental information disclosure: an empirical study of chinese listed companies. Journal of cleaner production 17, 593–600.

Malhotra, M.K., Grover, V., 1998. An assessment of survey research in pom: from constructs to theory. Journal of Operations Management 16, 407–425. URL: https://www.sciencedirect.com/science/article/pii/S0272696398000217, doi:https://doi.org/10.1016/S0272-6963(98)00021-7.

Martínez, E.O., 1990. La encuesta personal, postal y telefónica, in: Manual de investigación comercial, Pirámide. pp. 127–141.

Moktadir, M.A., Ahmadi, H.B., Sultana, R., Liou, J.J., Rezaei, J., et al., 2020. Circular economy practices in the leather industry: A practical step towards sustainable development. Journal of Cleaner Production 251, 119737.

Morera, J.M., Bacardit, A., Ollé, L., Bartolí, E., Borràs, M.D., 2007. Minimization of the environmental impact of chrome tanning: A new process with high chrome exhaustion. Chemosphere 69, 1728–1733.

Munasinghe, M., Shearer, W., 1995. Defining and measuring sustainability: the biogeophysical foundations. Technical Report. International Bank for Reconstruction and Development, Washington, DC . . . .

Nijkamp, P., Rodenburg, C.A., Verhoef, E.T., 1999. The adoption and diffusion of environmentally friendly technologies among firms. International Journal of Technology Management 17, 421–437.

Omoloso, O., Mortimer, K., Wise, W.R., Jraisat, L., 2020a. Sustainability research in the leather industry: a critical review of progress and opportunities for future research. Journal of Cleaner Production , 125441.

Omoloso, O., Wise, W.R., Mortimer, K., Jraisat, L., 2020b. Corporate sustainability disclosure: a leather industry perspective. Emerging Science Journal 4, 44–51.

Rao, J.R., Chandrababu, N., Muralidharan, C., Nair, B.U., Rao, P., Ramasami, T., 2003. Recouping the wastewater: a way forward for cleaner leather processing. Journal of cleaner production 11, 591–599.

Redwood, M., 2013. Corporate social responsibility and the carbon footprint of leather. J. Soc. Leather Technol. Chem. 97, 47–55.

Redwood, M., 2018. URL: https://www.leathernaturally.org/getattachment/Latest/Featured-Articles/Education/White-Paper-Explores-Sustainability-of-Leather/White_Paper_Sustainability_Leather_Naturally_Download.pdf.aspx?lang=en-US.

Restrepo-Collazos, S.P., Castaño-Martínez, C.A., 2021. Social responsibility: a look to the challenges faced by an university and a leather goods company, in bogota, colombia. Journal of Administrative Science 3, 29–32.

Sathish, M., Madhan, B., Sreeram, K.J., Rao, J.R., Nair, B.U., 2016. Alternative carrier medium for sustainable leather manufacturing–a review and perspective. Journal of cleaner production 112, 49–58.

Selvaraj, S., Jeevan, V., Jonnalagadda, R.R., Fathima, N.N., 2019. Conversion of tannery solid waste to sound absorbing nanofibrous materials: A road to sustainability. Journal of Cleaner Production 213, 375–383.

Shukla, P., Singh, J., Banerjee, M., 2015. They are not all same: variations in asian consumers' value perceptions of luxury brands. Marketing Letters 26, 265–278.

Siegle, L., 2011. To die for: is fashion wearing out the world? .

Śmiechowski, K., Lament, M., 2017. Impact of corporate social responsibility (csr) reporting on pro-ecological actions of tanneries. Journal of Cleaner Production 161, 991–999.

Spence, L.J., 1999. Does size matter? the state of the art in small business ethics. Business ethics: a European review 8, 163–174.

Sundar, V.J., Muralidharan, C., Mandal, A.B., 2013. A novel chrome tanning process for minimization of total dissolved solids and chromium in effluents. Journal of cleaner production 59, 239–244.

Sundar, V.J., Rao, J.R., Muralidharan, C., 2002. Cleaner chrome tanning—emerging options. Journal of cleaner production 10, 69–74.

Thanikaivelan, P., Rao, J.R., Nair, B.U., Ramasami, T., 2005. Recent trends in leather making: processes, problems, and pathways. Critical Reviews in Environmental Science and Technology 35, 37–79.

Thijssens, T., Bollen, L., Hassink, H., 2016. Managing sustainability reporting: many ways to publish exemplary reports. Journal of Cleaner Production 136, 86–101.

Vintro, C., Comajuncosa, J., 2010. Corporate social responsibility in the mining industry: criteria and indicators. Dyna 77, 31–41.

Vintró, C., Fortuny, J., Sanmiquel, L., Freijo, M., Edo, J., 2012. Is corporate social responsibility possible in the mining sector? evidence from catalan companies. Resources Policy 37, 118–125.

Vintró, C., Sanmiquel, L., Freijo, M., 2014. Environmental sustainability in the mining sector: evidence from catalan companies. Journal of cleaner production 84, 155–163.

WCED, S.W.S., 1987. World commission on environment and development. Our common future 17, 1–91.

Wilburn, K., Wilburn, R., 2013. Using global reporting initiative indicators for csr programs. Journal of Global responsibility .

Zuriaga-Agustí, E., Galiana-Aleixandre, M., Bes-Piá, A., Mendoza-Roca, J., Risueño-Puchades, V., Segarra, V., 2015. Pollution reduction in an eco-friendly chrome-free tanning and evaluation of the biodegradation by composting of the tanned leather wastes. Journal of Cleaner Production 87, 874–881.

Computational Economics
https://doi.org/10.1007/s10614-021-10208-4

An Application of the IFM Method for the Risk Assessment of Financial Instruments

Adrià Pons [1] · Eduard Cristobal-Fransi [1] · Carla Vintrò [1] · Josep Rius[1] · Oriol Querol [2] · Jordi Vilaplana[2]

Accepted: 23 September 2021

Abstract

External influences or behavioral biases can affect the way risk is perceived. This paper studies the prediction of VaR (Value at Risk) as a measure of the risk of loss for investments on financial products. Our aim is to predict the percentage of loss that a financial product would have in the future to assess the risks and determine the potential loss of a security in the stock market, thus reducing reasoning influenced by feelings for bank and financial firms seeking to deploy AI and advanced automation. We used the IFM (inference function for margins) method in different market scenarios, with particular emphasis on the strengths and weaknesses of it. The study is assessed on single product level with the skewed studen-t GARCH(1,1) model and portfolio level with t-copulas for the inter-dependencies. It has been shown that under normal market conditions the risk is predicted properly for both levels. However, when an unexpected market event occurs, the prediction fails. To address this limitation, a combined model with sentiment analysis and regression is proposed for further investigation as a future work.

Keywords Risk simulation · Monte carlo · GARCH · t-Copula · VaR · Risk tolerance · Behavioral finance · Smart banking

1 Introduction

Customer experience has emerged as a new battleground in investment management. AI is changing how financial institutions attract and retain customers, and through this, offers the opportunity for firms to innovate and enhance the investor

✉ Josep Rius
josep.riustorrento@udl.cat

[1] Department of Business Management, University of Lleida, C/Jaume II, 73, 25001 Lleida, Spain

[2] Department of Computer Science, University of Lleida, C/Jaume II, 69, 25001 Lleida, Spain

Published online: 13 October 2021

journey. What is certain now is that investment management firms can no longer rely solely on price and outperformance to attract investors. Firms that adapt their apps and integrate AI, data, and analytics into their bank solutions will be better placed to optimize and execute their product and content distribution strategies.

On the other hand, human decisions are largely relied upon on how information is represented by third parties. According to Pompian (2017), the way investors think and feel affects their investment behaviors which is unconsciously influenced by past experiences and personal beliefs to the extent that even intelligent investors may deviate from logic and reason. Bollen et al. (2011) find that Twitter mood predicts subsequent stock market movements. Gilbert and Karahalios (2010) find that the level of anxiety of posts on the blog site Live Journal predicts price declines. Behavioral finance suggests that the investment decision-making process is influenced by various behavioral biases that encourage investors to deviate from rationality and make irrational investment decisions (Kumar and Goyal, 2015).

Investors' perceptions regarding the risk and return characteristics of a particular stock or the stock market are commonly assumed to be key drivers of their decision making (McInish and Srivastava, 1984; Antonides and der Sar, 1990). This means that investors must decide which risks to take and how much to take.

While previous experiments have already shown that emotions can increase risk aversion (Kuhnen and Knutson, 2005, 2011; Knutson et al. 2008), our goal is to provide some prior knowledge of the likely risk scenarios that can face different type of investors (i.e. floor brokers, in-house traders, institutional traders), and to provide an algorithmic trading system for financial enterprises that intend to take full advantage of AI applying machine learning techniques to more accurately move its smart banking solutions from manual to semi-automated or fully automated processes and help their customers segments which have been traditionally underserved to become more rational and unbiased using the IFM method (inference function for margins) proposed in Xu (1996) and which is widely used on the financial industry.

Precisely we have selected the assumption of skewed student-t distributed residuals on a GARCH(1,1) regression and copulas for the inter-dependencies, implemented by one of the biggest German banks (for disclosure contract, the name of the bank is confidential) because of its reliability. Models, whose only input are the historical prices of the forecasted securities.

We believe that this method is not sufficient in certain market situations. The main purpose is the testing of the IFM method in different scenarios, where each scenario is composed of two simulations in two consecutive time intervals in the past (in normal market conditions and in specific market events that changed the course of securities). To assess the perpetuation of the risk extracted from both simulations.

The risk measure used is the value at risk (Linsmeier and Pearson, 1996; Jorion, 1997), which estimates how much might a set of investments lose (with a given probability), in a set time period such a year. A Monte Carlo Simulation (Glasserman, 2003) with a GARCH process is performed in order to calculate the expected return, volatility and the value at risk (Wiener and Benninga, 1998), on both, portfolio and instrument level (Gueant 2012).

The project gives an insight of the current research on the topic, followed by a detailed explanation of the method, the data-set used for the simulation, the results obtained, the discussion of the results and finally the conclusions for the future research of a new combined model.

2 Theoretical Background

Traditional financial theories consider that investors operate rationally in making financial decisions (Kiymaz et al., 2016) and evaluate possible alternatives on the basis of utility and associated risk. Risk is understood as the degree of uncertainty or potential financial loss inherent in an investment decision. A preliminary concept was introduced in Modern Portfolio Theory by Nobel laureate Harry Markowitz in his paper "Portfolio Selection" published in 1952 by the Journal of Finance.

However, different studies reported that investors usually do not make rational or logical decisions. Under this context, in the 1980s, behavioral finance emerged as a new concept in the fields of economics and finance. It studies the psychological and behavioral aspects, and the irrationality of investors in economic and financial decision-making (Barber and Odean, 2000; Weisbenner and Ivkovich, 2003; Statman et al., 2006). Different authors have conducted studies in this field. For instance, M. Barber and Odean (2001) studied behavioral changes in individual investment decisions. Kumar and Goyal (2015) analyzed how behavioral biases may influence the investor's rationality in investment decision-making even though risk is a determining factor when it comes to invest (McInish and Srivastava 1984). Bailey et al. (2011) show the effect of behavioral biases on mutual fund options. The results show that investors tend to make poor decisions about their investments. And Kiymaz et al. (2016), studied the behavioral biases of financial professionals in Turkey. These authors concluded that younger professionals, with less training, with less risk aversion, and with unique brokerage accounts are more likely to invest in stocks.

Another line of work focuses on trying to explain the behavior of investors using various dimensions in addition to the biases of investors. For example, Georgarakos and Fürth (2015) use financial competence and show that financial advice is more important for investors with little perceived financial competition. Hoffmann et al. (2013) state that investors who use fundamental analysis are more likely to take risks, have high trading volumes and are overconfident. van Rooij et al. (2011), on the other hand, conclude that basic financial education is positively related to participation in the stock market. Also, Nicolosi et al. (2009) say that investors learn from their investment experiences, despite presenting irrational behavior. Finally, comment that many studies use gender, marital status, impact of coworkers, financial education and cultural differences to explain the behavior of the investor (Bernard et al., 2018; Jacobsen et al., 2014; Halko et al., 2012; Heimer, 2014; Mugerman et al., 2014; Tekçe et al., 2016).

Researchers and financial institutions have created and applied models to assess the risk of financial products. Being the regression models, like the IFM method, first introduced by Joe and Xu (1996) the most used in the financial sector. These models take no more input than the time series of the securities, that is, sequence of prices in the market of a given security, in chronological order. The IFM method consists of doing separate optimization of the uni-variate likelihoods, followed by an optimization of the multivariate likelihood as a function of the dependence parameter vector. Further information regarding the IFM method can be found in McNeil et al. (2005), Nelder and Mead (1965).

The optimization of the uni-variate likelihoods, is given by maximizing the quasi maximum likelihood function (QMLE) instead of the Maximum likelihood, see Bollerslev and Wooldridge (1992). Where each d estimations follow the skewed-t GARCH(1,1) models (Bollerslev, 1986; Alberga et al., 2008; Jondeau et al., 2007; Lambert and Laurent, 2001; Pantelidis and Pittis, 2005). Whereas the multivariate likelihood, is on this case given by the ML of a d dimensional Student-t Copula (Embrechts et al., 2001; Nelsen, 2006; Scherer and Mai, 2014; Zhang and Ng, 2010), optimized by the BFGS algorithm. Finally the value at risk is forecasted with a Monte Carlo Simulation (Glasserman, 2003) following a GARCH(1,1) regression.

The method exposed on the next section follows the state of the art in one of the biggest German financial institutions. This method has been used in this work for the calculation of the results.

3 Methods

Our approach fits n GARCH(1-1) models in order to estimate the unconditional variance and the skewed studen-t distributed residuals on single product level. Next to a student-t copula fitting of n dimensions for the inter-dependencies on portfolio level. Finally, performing Monte Carlo simulations with the previous fitted models to estimate the future prices, used to calculate the VaR.

3.1 GARCH Fitting

Giving n observations of d individual financial products we denote the vector containing the n observations for an individual product j with Y_j, for $j \in 1, \ldots, d$. Hence Y_j is a $(1 \times n)$ vector for each j. Along these lines the matrix $Y = (Y_1, \ldots, Y_d)$ is $d \times n$. The standard GARCH(p, q) model estimates the variance of returns as a simple quadratic form, hence the variance can be expressed at time t as in Bollerslev (1986) by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Since we are following the IFM method, one is considering each financial product separately and therefore the fitting boils down to a GARCH(1, 1) model without drift $\mu = 0$ which is then given by

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{1}$$

The input parameters for the model are as follows:

- $\epsilon_t = y_{t,j}$, thereby $y_{t,j}$ denotes the (log) return of the j-th financial product at time $t = i$ (the i-th observation).
- $\epsilon_t = z_t\sigma_t$, where $z_t \sim st(0, 1, v_t, \lambda_t)$.

The skewed Student t-distribution (st) with mean zero and unit variance extends the regular Student t-density with $v \in (2, \infty)$ degrees of freedom by an additional skewness parameter $-1 < \lambda < 1$. The density f for z_t is defined as (Hansen, 1994):

$$f(z_t, 0, 1, v, \lambda) = \begin{cases} bc\left(1 + \frac{1}{v-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(v+1)/2} & for \quad z_t < -\frac{a}{b} \\ bc\left(1 + \frac{1}{v-2}\left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(v+1)/2} & for \quad z_t \geq -\frac{a}{b} \end{cases} \tag{2}$$

thereby the constants a, b and c are defined by

$$a = 4\lambda c\left(\frac{v-2}{v-1}\right), \tag{3}$$

$$b^2 = 1 + 3\lambda^2 - a^2, \tag{4}$$

$$c = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)}\Gamma(v/2)} \tag{5}$$

An alternative definition of the density function f can be found in Lambert and Laurent (2001). The next step is to express the ML function in order to obtain the residuals ϵ and the optimal parameter vector , i.e $\Theta = (v, \lambda, \sigma_t)$, i.e $\Theta = (v, \lambda, \omega, \alpha, \beta)$. Instead of maximizing the ML function, it can be also optimized the quasi maximum likelihood function (QMLE) see Bollerslev and Wooldridge (1992)

$$L_T(z_1, \ldots, z_T; \Theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\Theta) = \frac{1}{T}\sum_{t=1}^{T} \log\left(\frac{f(z_t; v, \lambda)}{\sigma_t}\right). \tag{6}$$

Thereby $t = 1, \ldots, T$ denotes the timestamp of the transformed observations $z_t = \frac{\epsilon_t}{\sigma_t}$. We can also write $t = 1, \ldots, T$, $i = 1, \ldots, n$ for n observations collected for the specific financial product. After optimizing the ML function obtains $\hat{\Theta}$ and with the optimal values for σ_t for $t = 1, \ldots, T$ and equation (2.1) we get the optimal values for α, β and ω. To obtain the optimal values for v, λ and σ we use the direct search method by Nelder and Mead (1965) (downhill simplex). By minimizing the negative QMLE function in (2.6) in combination with the formula (2.4) we obtain the volatility scaling parameters, the distribution parameters and the residuals (marginal distribution parameters for the Copula estimation).

3.2 Copula Fitting

In this section it is described how to fit a multivariate t-copula to a set of d financial products i.e returns using Kendalls Tau for estimating ρ. The parameter for the degrees of freedom v is obtained via a ML estimation. The density of a d-dimensional Student-t-Copula with parameters ρ and v degrees of freedom can be written as follows (Demarta and McNeil, 2004):

$$c^t_{\rho,v}(u_1,\ldots,u_d) = \frac{f_{\rho,v}(t_v^{-1}(u_1),\ldots,t_v^{-1}(u_d))}{\prod_{j=1}^{d} f_v(t_v^{-1}(u_i))} \tag{7}$$

Thereby $t_v^{-1}(u_1)$ represents the corresponding quantile of a student t distribution with v degrees of freedom. The function $f_{\rho,v}$ defines the density of a multivariate student-t distribution and along these lines f_v the density of a univariate student-t distribution. The nominator in (7) can be written as follows (Demarta and McNeil, 2004):

$$f_{\rho,v}(t_v^{-1}(u_1),\ldots,t_v^{-1}(u_d)) = \frac{\Gamma\frac{v+d}{2}}{\Gamma(\frac{v}{2})v^{d/2}\pi^{d/2}\rho^{1/2}[1+\frac{1}{v}(q-\mu)'\rho^{-1}(q-\mu)]^{(v+d)/2}} \tag{8}$$

The vector $q = (q_1,\ldots,q_d)$ denotes the quantile of a Student t distribution for the ith observation of the jth risk factor after a strictly increasing transformation. Hence for one point in time (fix i), q maps all risk factors $(u1,\ldots,ud)$ to their corresponding quantiles with values in $\mathbb{R}$. Hence for each i (which can be interpreted as the time) we obtain a d dimensional vector q. For $d = 1$ we obtain the uni variate Student's t-distribution, which looks as follows

$$f_v(t_v^{-1}(u_j)) = \frac{\Gamma\frac{v+1}{2}}{\Gamma(\frac{v}{2})\sqrt{v\pi}(1+\frac{q_j^2}{v})^{(v+1)/2}} \tag{9}$$

Hence the density of a d dimensional Student t Copula with parameters ρ and v degrees of freedom can be written as follows

$$c^t_{\rho,v}(u_1,\ldots,u_d) = \frac{1}{\sqrt{|\rho|}}\frac{\Gamma(\frac{v+d}{2})\Gamma(\frac{v}{2})^{d-1}}{\Gamma(\frac{v+1}{2})^d}\frac{\prod_{j=1}^{d}(1+\frac{q_j^2}{v})^{\frac{v+1}{2}}}{(1+\frac{q'\rho^{-1}q}{v})^{\frac{v+d}{2}}} \tag{10}$$

Thereby $|\rho|$ denotes the determinant of the input matrix ρ and $\Gamma(\cdot)$ the gamma function. Since we are interested in the optimal parameters for ρ and v within the next subsections of this manual the necessary steps and mathematical derivations are provided, derived and explained in more detail.

3.3 The Rank Transformation

We assume that the residual vector $\epsilon_1,\ldots,\epsilon_d$ is given, whereas each vector is assumed to have a different distribution F_j for $j = 1,\ldots,d$ and $i = 1,\ldots,n$

observations. The approach follows Scherer and Mai (2014) and works as follows: For each time series Y_j the idea is to replace the smallest of these number with the value $\frac{1}{n+1}$, the second smallest with the value $\frac{2}{n+1}$ and so on. In general terms, the data is transformed in such a way that no information regarding the dependence is lost. Further it holds true that now all values are within the unit interval $[0, 1]$. Since copulas are defined for values in $[0, 1]$ this step is crucial in order to transform the returns ϵ to U_j.

3.4 Estimating ρ

The method for estimating ρ is oriented on McNeil et al. (2005) p.231. Hence calibrating t copulas using Kendalls Tau τ (for a formal definition of Kendalls Tau see Nelsen (2006) p.158). Given a vector U the relationship between Kendalls Tau and the correlation parameter of the student t copula $C_t(v, \rho)$ is given by

$$\rho_\tau(U_i, U_j) = \frac{2}{\pi} \arcsin(\rho_{i,j}) \tag{11}$$

Since one can directly calculate Kendalls Tau (the left hand side) we can invert (11) in order to obtain the following estimator for the copula parameter ρ.

$$\rho_\tau(U_i, U_j) = \frac{2}{\pi} \arcsin(\rho_{i,j}) \Rightarrow \sin\left(\rho_\tau(U_i, U_j)\frac{\pi}{2}\right) = \rho_{i,j}. \tag{12}$$

Hence the estimated symmetric matrix $\hat{\rho}$ which is dxd has the following shape:

$$\hat{\rho} = \begin{bmatrix} 1 & \sin(\rho_\tau(U_1, U_2)\frac{\pi}{2}) & \cdots & \sin(\rho_\tau(U_1, U_d)\frac{\pi}{2}) \\ \sin(\rho_\tau(U_2, U_1)\frac{\pi}{2}) & 1 & \cdots & \sin(\rho_\tau(U_2, U_d)\frac{\pi}{2}) \\ \vdots & \vdots & \ddots & \vdots \\ \sin(\rho_\tau(U_d, U_1)\frac{\pi}{2}) & \sin(\rho_\tau(U_d, U_2)\frac{\pi}{2}) & \cdots & 1 \end{bmatrix} \tag{13}$$

The easiest way to obtain the desired matrix $\hat{\rho}$ is to calculate Kendalls Tau empirically for each pair (U_i, U_j) followed by transforming each entry of the matrix with the given sinus transformation. Once calculated, $\hat{\rho}$ will be used as starting point for the optimizer of the log likelihood function after applying a QR decomposition of the matrix.

3.5 Maximum Likelihood Estimation

The estimation of the empirical ρ in section 4.4 can be seen as a starting point for the further optimization. A more sophisticated approach for finding the relevant parameters is the simultaneously estimation of ρ and v. The necessary input vector is given by the transformed values, denoted by u. Due to the fact that $\log(\prod_i a_i) = \sum_i \log(ai)$ and $\log(\frac{a}{b}) = \log(a) - \log(b)$ one can rewrite the ML function for the density in (7) and a given observation as follows

$$\begin{aligned}\log(c^t_{\nu,\hat{\rho}}(u_1,\ldots,u_d)&=\log\left(\frac{f_{\nu,\rho}(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_d))}{\prod_{j=1}^d f_\nu(t_\nu^{-1}(u_j))}\right)\\&=\log\bigl(f_{\nu,\rho}(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_d))\bigr)-\log\left(\prod_{j=1}^d f_\nu(t_\nu^{-1}(u_j))\right)\\&=\log\bigl(f_{\nu,\rho}(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_d))\bigr)-\sum_{j=1}^d\log(f_\nu(t_\nu^{-1}(u_j)))\end{aligned}\tag{14}$$

3.6 Broyden–Fletcher–Goldfarb–Shanno Algorithm

In order to solve the stated ML equations we are applying the Broyden–Fletcher–Goldfarb–Shanno algorithm which is iterative method for solving unconstrained non linear optimization problems. The algorithm itself belongs to the class of Quasi-Newton method, which is characterized by the fact that the algorithm is approximating the Hessian matrix.

3.7 Monte Carlo Simulation

The standard Monte Carlo approach exploits the Law of Large Numbers Theorem in order to get an estimation for a theoretical expectation, such as the price of a financial product. With the Monte Carlo path simulations we are able to compute a theoretical expectation numerically by constantly repeating a specific simulation and averaging the obtained result. These random numbers will be used to estimate the return of the asset at the end of the analysis horizon. N paths of independent standardized residuals over m days horizon are going to be obtained in order to calculate the expected return, volatility and the value at risk (VaR). This method is especially useful if the expectation that we are interested in has no closed analytic solution. Another advantage is that the Monte Carlo estimator is easy to implement and works independently of the underlying distribution and hence for a tremendous class of stochastic processes.

3.7.1 Monte Carlo Simulations Including GARCH Models and the t-Copula

According to Bollerslev (1986), given the GARCH(1,1) model (Eq. 1), the optimal predictor $\hat{\sigma}^2_{t+s}$ of the conditional variance for forecast horizon s is the conditional expected value:

$$\hat{\sigma}^2_{t+s}=E_t[\sigma^2_{t+s}]=\omega\sum(\alpha+\beta)^{i-1}+(\alpha+\beta)^{s-1}\sigma^2_{t+1'}\tag{15}$$

where ω, α, β and σ_{t+1} are obtained through the GARCH fitting process. Moreover, as $\alpha+\beta<1$ the unconditional variance can be expressed:

$$\sigma^2 = \mathrm{Var}(\epsilon_t) = \frac{\omega}{1-\alpha-\beta} \tag{16}$$

in that case, (15), can be written as follows:

$$\hat{\sigma}^2_{t+s} = \sigma^2 + (a+b)^{s-1}(\sigma^2_{t+1} - \sigma^2). \tag{17}$$

Since $\alpha + \beta < 1$, $(a+b)^{s-1}$ converges to zero for $s \to \infty$ and therefore $\hat{\sigma}^2_{t+s} \to \sigma^2$ for $\to \infty$. Hence the predictor tends to the unconditional variance.

The innovation process used to forecast the returns is described by the following formula

$$\epsilon_t = z_t \sigma_t + \mu \tag{18}$$

where z_t is the random generated residual and μ denotes the drift factor.

3.7.2 Simulations on a Single Product Level

In the case of independent products, z_t follows a skewed student-t distribution (2) for the innovations. Hence $z_t \sim st(0, 1, v_t, \lambda_t)$ is generated through the inverse transform sampling method, i.e. it is obtained by the quantile function which is defined as follows (Jondeau et al. 2007):

$$F^{-1}(y) = \begin{cases} \frac{1}{b}[(1-\lambda)\sqrt{\frac{v-2}{v}}T^{-1}(\frac{y}{1-\lambda}|v) - a] & for \quad y < \frac{1-\lambda}{2} \\ \frac{1}{b}[(1+\lambda)\sqrt{\frac{v-2}{v}}T^{-1}(\frac{y}{1-\lambda}|v) - a] & for \quad y \geq \frac{1-\lambda}{2} \end{cases} \tag{19}$$

where $T(x|v)$ is the cdf of the standard t-distribution with v degrees of freedom and can be expressed with (9)

$$T(x|v) = \int_{-\infty}^{x} f_v(w)dw = \int_{-\infty}^{x} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}(1+\frac{w^2}{v})^{(v+1)/2}} \tag{20}$$

3.7.3 Simulations Considering Interdependencies in the Portfolio

Due to the fact that the d multivariate student-t distribution can be written as follows

$$Z \stackrel{d}{=} \mu + \frac{\sqrt{v}}{\sqrt{S}} X \tag{21}$$

with $\mu \in \mathbb{R}^d, S \sim x^2_v, X \sim N(0, \Sigma)$ we can easily adapt the simulation algorithm as described in Embrechts et al. (2001):

(1) Calculate the Cholesky decomposition A of $\hat{\rho}$.
(2) Simulate d independent random variates $x_1, \ldots, x_n$ from $N(0, 1)$
(3) Simulate a random variate s from χ^2_v independent of $x_1, \ldots, x_n$.

(4) Set $y = Ax$
(5) Set $z = \frac{\sqrt{v}}{\sqrt{s}} y$
(6) Set $u_i = t_v(z_i)$, for $i = 1, \ldots, d$

Doing so, it follows that $(u_1, \ldots, u_n) \sim C^t_{v,\hat{\rho}}$.

Finally, given the simulated returns of each index, the portfolio weights are applied to the respective simulated return paths and the joint portfolio return is calculated by summing up the instrument specific weighted return simulations.

$$\mathbb{E}\left[\sum_{i=1}^{d} \omega_i R_t\right] = \sum_{i=1}^{d} \mathbb{E}[\omega_i R_t] = \sum_{i=1}^{d} \omega_i \mathbb{E}[R_t]. \tag{22}$$

Through the linearity of the expectation, the portfolio value at a specific date can be easily traced back to the Monte Carlo simulation of the d individual assets.

For the Monte Carlo Simulation we simulate each product over a time horizon of 252 days. Hence there are $252 \cdot d$ simulations for the d individual products. The Monte Carlo simulation takes place by repeating this procedure $N = 1000$ times.

3.7.4 Expected Volatility

Due to the Continuous Mapping Theorem we can additionally obtain the following even stronger and more useful result: For a function f with $\mathbb{E}[|f(X)|] < \infty$ it holds that

$$\frac{1}{N} \sum_{i=1}^{N} f(X_i) \xrightarrow{a.s} \mathbb{E}[f(X)]. \tag{23}$$

With the Continuous Mapping Theorem (23) we can calculate the deviation of the expected returns. The volatility is measured by the standard deviation of the N forecasted returns x_i, where μ denotes the sample average:

$$\begin{aligned} \mathbb{E}[(R_t - \mathbb{E}[R_t])^2] := \hat{\sigma}^2 &= \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{\mu})^2 \\ \Rightarrow \hat{\sigma} &= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{\mu})^2} \end{aligned} \tag{24}$$

3.7.5 Value at Risk

The Value at Risk calculates the largest loss likely to be suffered on an investment (portfolio or product) with a given probability, over a holding period of time. Given a confidence level $\alpha \in (0, 1)$, the VaR of the distribution formed by the n forecasted returns, is calculated by taking the smallest return that exceeds the probability $1 - \alpha$.

After the simulation, N Portfolio values over the 252 days horizon have been generated by applying (22). The distribution followed by these paths allows to calculate the VaR at a given confidence level.

4 Data

The main purpose of this research is to calculate the risk of four financial products (BMW, Tesla, Samsung and Facebook) considering two different scenarios (with and without potentially favorable/unfavorable events) and level (low/high) of return and volatility. The data used for each scenario consist of three years of daily closing prices (ca. 750 observations, extracted from Thomson Reuters) starting in two different points of time. This three years period will be enough to perform the Fitted Garch model. On the other hand, the interval between the two points will be a fixed estimation window of 365 days, which is equal to the horizon length in the Monte Carlo simulation.

First case Financial product with low average return. BAYERISCHE MOTOREN WERKE AG (DE0005190003). First data set range 25.08.14–24.08.17, second data set range 24.08.15–24.08.18 (see Fig. 1).

Second case Financial product with high return and high volatility. Tesla Motors (US88160R1014). First data set range 25.08.14–24.08.17, second data set range 24.08.15–24.08.18 (see Fig. 2).

Third case Financial product with high return and low volatility. Samsung Electronics (US7960508882). First data set range, 13.02.08–14.02.11 second data set range 13.02.09–13.02.12 (see Fig. 3). The end of the first data set coincides with the release of the Samsung Galaxy S2, a positive event that raised the price of the financial instrument.

Forth case Financial product with high return and high volatility. Facebook (US30303M1027). First data set range 25.08.14–24.08.17, second data set range 24.08.15–24.08.18 (see Fig. 4). The end of the first data set coincides with legal accusations and controversy regarding the data protection of the platform users, a negative event that dropped the price of the financial instrument.

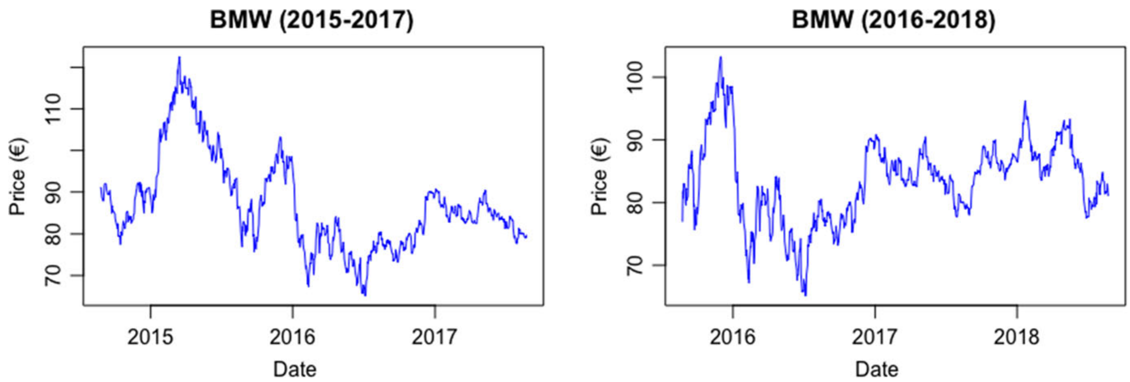


Fig. 1 Financial product with low average return (BMW)

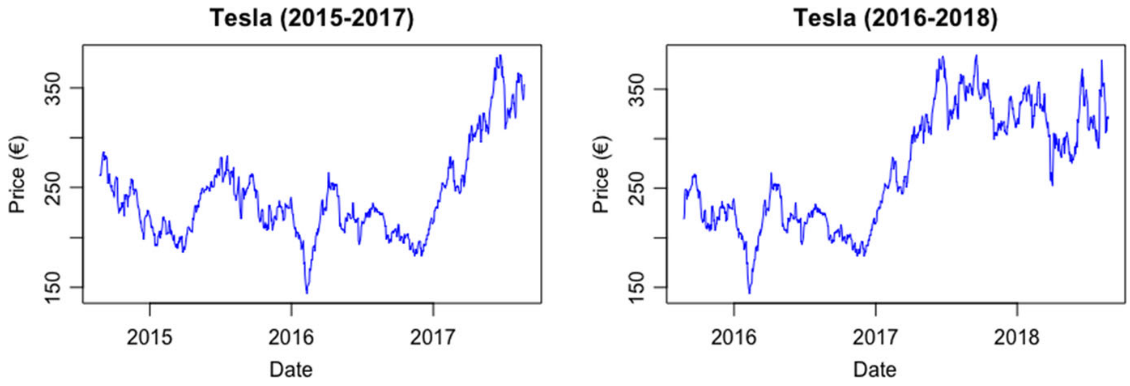


Fig. 2 Financial product with high return and high volatility (Tesla Motors)

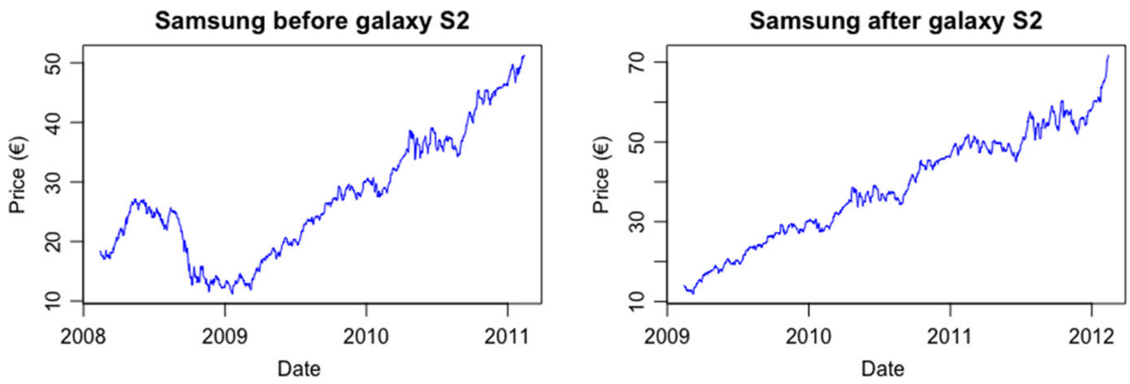


Fig. 3 Financial product with high return and low volatility (Samsung)

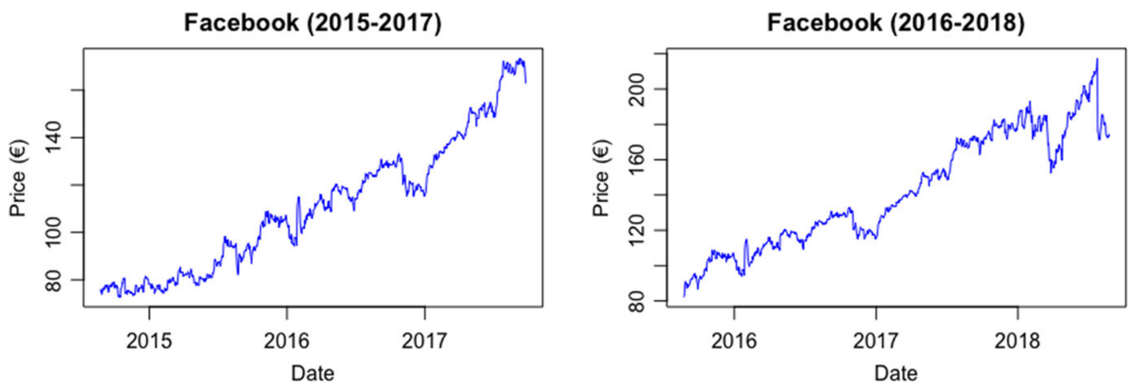


Fig. 4 Financial product with high return and high volatility (Facebook)

5 Results

The log returns (Fig. 5) are the input parameters for the GARCH models. From the results above it can be noticed that Tesla and BMW haven't got any observable difference between the two intervals (having BMW lower peaks than Tesla, which corresponds to the lower return of the first one).

However, on the other side, it can be observed that Samsung's volatility decreases on the second interval due to the fact that the 2008 drop is not taken into account.

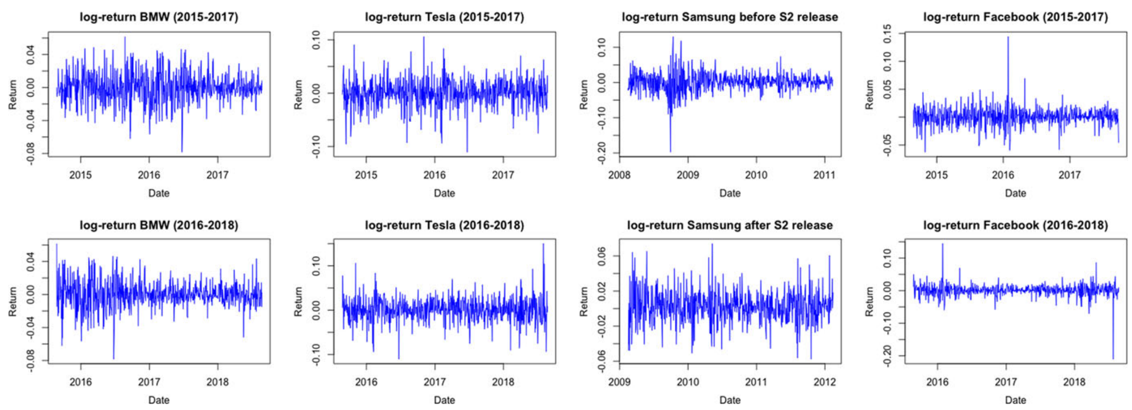


Fig. 5 Input parameters for GARCH models (BMW, Tesla, Samsung, Facebook)

Finally, Facebook volatility increases on the second interval because of the huge change of prices at the middle of 2018.

The parameters omega, alpha, beta, skew and shape (Table 1) are used to calculate the unconditional variance, the deviation and the random generated residuals in both simulations: single product and considering inter-dependencies.

It can be appreciated form the previous table that BMW is the only financial product positively skewed, which implies that the majority of the samples fall towards the lower side.

Additionally, Tesla and Facebook obtained a lower shape, which will create distributions with fatter tails, which will create more extreme random values on the forecast.

5.1 Simulation on a Single Product Level

Figure 6 shows the different Monte Carlo simulation stages for the first case study, BMW, period 2015–2017 [with 1 path, 10 paths, 50 paths and 100 paths consecutively]. To calculate the VaR, 1000 paths are used.

After the 1000 simulations, we can visually confirm that the results (Fig. 7) follow skewed student-t distributions. BMW slightly positively skewed (really close

Table 1 Parameters for single product simulation and considering inter-dependencies

Security	Omega	Alpha	Beta	Skew	Shape
BMW 2015–2017	$5.0401e-07$	$3.2676e-02$	$9.6518e-01$	$1.0749e+00$	$5.5536e+00$
BMW 2016–2018	$3.0342e-10$	$3.3622e-02$	$9.6703e-01$	$1.0179e+00$	$6.8625e+00$
Tesla 2015–2017	$1.4394e-05$	$1.8666e-02$	$9.5952e-01$	$9.2496e-01$	$4.5832e+00$
Tesla 2016–2018	$1.1827e-05$	$2.6065e-02$	$9.5804e-01$	$9.6458e-01$	$4.8743e+00$
Samsung before s2	$3.5777e-06$	$3.2676e-02$	$9.2256e-01$	$9.3432e-01$	$7.1132e+00$
Samsung after s2	$1.2977e-05$	$7.7391e-02$	$8.8578e-01$	$9.4029e-01$	$6.1430e+00$
Facebook 2015–17	$6.2301e-06$	$7.6749e-02$	$9.0399e-01$	$9.1626e-01$	$4.3030e+00$
Facebook 2016–18	$2.1480e-05$	$1.3733e-01$	$8.1097e-01$	$9.1075e-01$	$3.3055e+00$

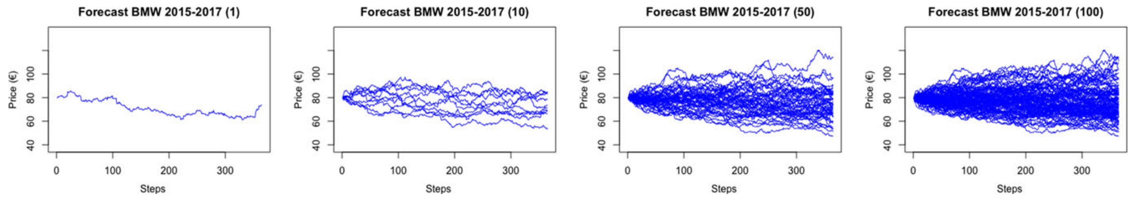


Fig. 6 Stages of a Monte Carlo simulation for BMW

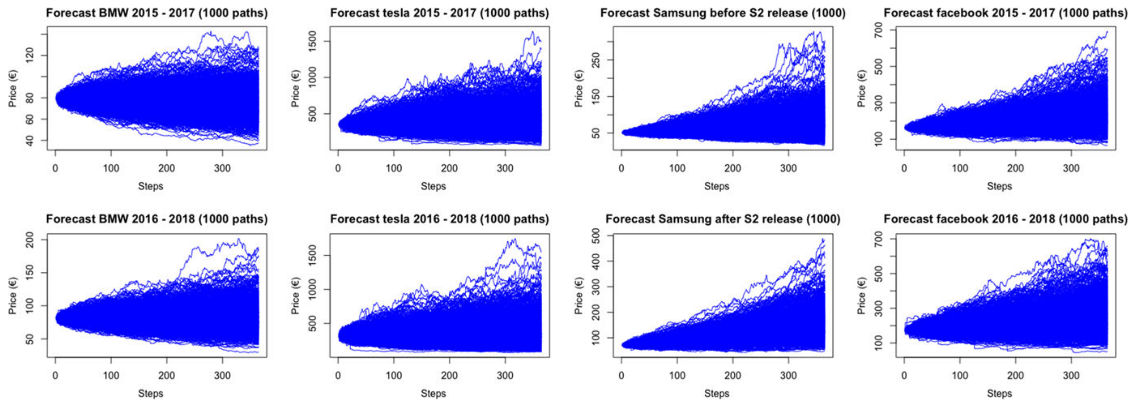


Fig. 7 1000 paths resulting from the Monte Carlo simulation

to a studen-t distribution without skew on the 2016–2018 data set due to the lower skew parameter) and the other three instruments skewed negatively.

Figure 8 shows the end prices of each of the 1000 paths. The value at risk price is highlighted on red (0,05 quantile of the distribution) and the threshold representing the initial price. The bigger the difference between the VaR price and the threshold, the bigger the loss could be on a worst market scenario.

It can easily be noticed that this difference is quite low on the Samsung forecasts (third case study), or quite big on the Tesla case (second case study). On the other side, Tesla, Samsung or Facebook would perform really well on good market scenarios

The VaR(%) on BMW and Tesla before and after the 365 day interval slightly changes (Table 2). This was already noticed on the log-returns of Fig. 5, where the

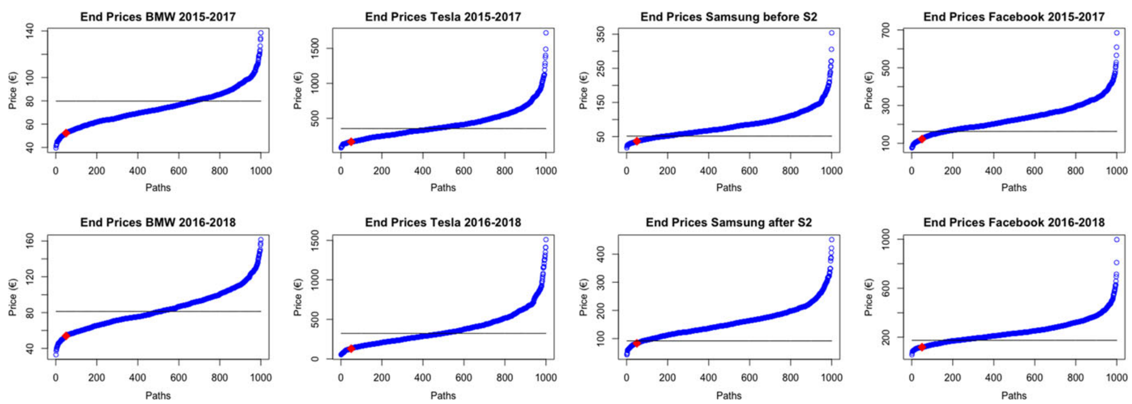


Fig. 8 Ordered end prices for the 1000 paths (blue points), VaR price (red point) and initial price treshhold (black line)

Table 2 VaR before and after the 365 day interval for all four financial products

VaR(%)	BMW	Tesla	Facebook	Samsung
Before	−33, .2	−55.89	−20.20	−28.25
After	−33.34	−58.81	−33.03	−5.01

two data sets had very similar graphs on both dates. Also, in a worst market scenario, Tesla could lose 55% of its current value, in comparison to BMW, 33%, which makes it the riskiest product among the four.

On the other hand, Facebook and Samsung register a change on their risk after the interval. The first one gaining a 13% more risk and the second one by lowering to only 5%. Samsung's case was already spotted on the log returns graph as the low price spike was not present on the second interval.

5.2 Simulations Considering Interdependencies in the Portfolio

Figure 9 shows the bi-variate copulas created from the combination of the three securities. It can be observed that the only two financial products that are slightly correlated are BMW and Tesla (first and second case study). As it can be observed on the graph, the data points are grouped towards the diagonal formed between the point (0,0) and (1,1). Whereas in the other combinations, the points are not dispersed around the space. However, the results are calculated for a portfolio that contains the three securities at the same time, which is translated as a t-student multivariate copula of 3 dimensions, formed by the uniform ranked residuals obtained of the garch fitting models:

In the Table 3 it can be noticed that the data follows rather a normal distribution than the student-t, as the degrees of freedom are really big. Additionally, rho is closer to 0 which emphasizes that there is nearly no correlation between the three securities.

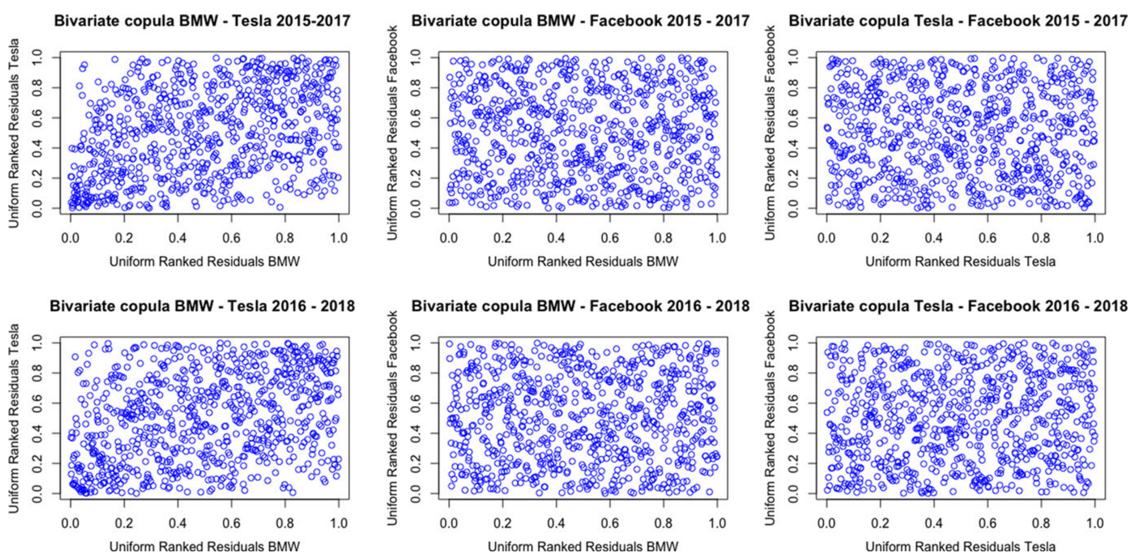


Fig. 9 Bi-variate copulas between the BMW-Tesla, BMW-Facebook and Tesla-Facebook

Table 3 Fitted multivariate copula models (3 dimensions)

Copula	Rho	Shape
Before	0.1138859	36.0797850
After	0.1434166	32.3817758

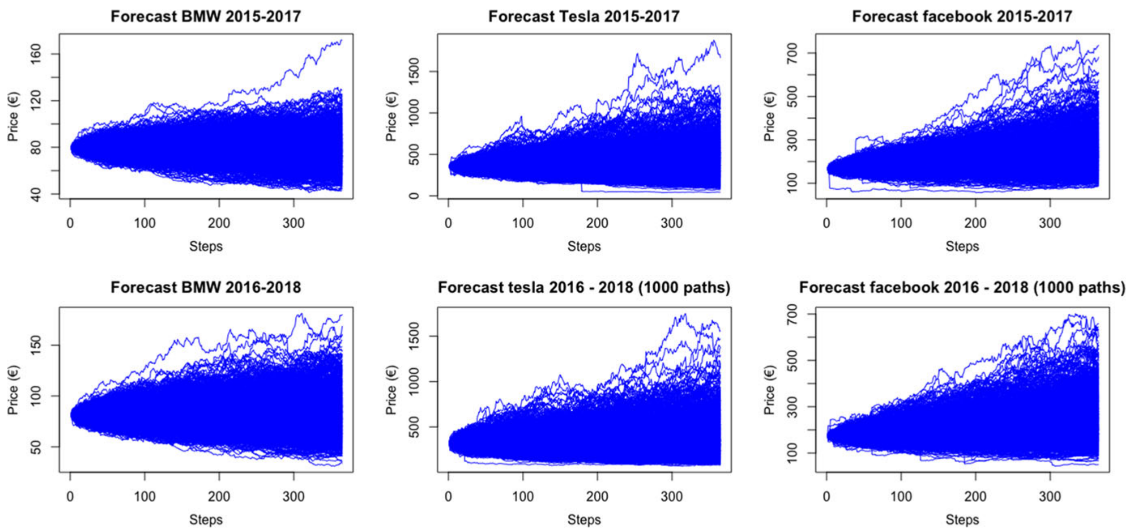


Fig. 10 1000 paths resulting from the Monte Carlo simulation including interdependencies

After the 1000 simulations (Fig. 10), as in the case of single products, it can visually be confirmed that the results coming from the multivariate random generator also follow skewed student-t distributions.

As showed on the Fig. 9, the securities BMW and Tesla share a certain correlation. This is also confirmed on the end prices (Fig. 11), as taking into account inter-dependencies BMW would perform much better on a good market scenario (Tesla does so on single product simulations). On the opposite side, Facebook does

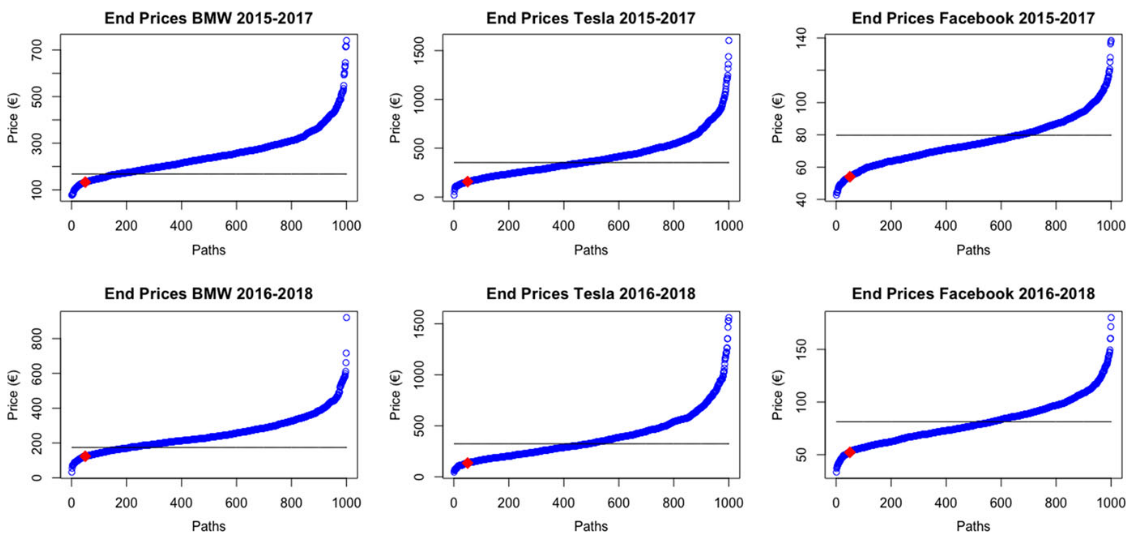


Fig. 11 Ordered end prices for the 1000 paths (blue points), VaR price (red point) and initial price treshhold (black line) including interdependencies

not perform so well on bad market scenarios due to the influence of the two other financial products.

Comparing the VaR(%) with the results from the simulations on a single product level (Table 4), we can observe that the good results of Facebook before the interval smooths the risk of BMW (from 33% to 29%) and Tesla (from 55% to 53%). However, Facebook gets penalized by the two other securities (from 20% to 26%). After the interval, BMW's risk gets increased drastically in comparison to single product simulations(33% to 37%) whether on the opposite Facebook is getting a lower score (33% to 30%).

Assuming that the weight of the three securities is the same 0.33%, four portfolios with three securities (BMW, Tesla and Facebook) have been created and it's overall VaR(%) calculated (Table 5).

As it can be appreciated, the overall Risk of the portfolio doesn't barely change during the same periods whether the risk has been calculated through single products or considering inter-dependencies. Nevertheless, the risk in the interdependent portfolio is slightly bigger due to the fact that two securities (BMW and Tesla) are correlated.

6 Discussion

The results show that the Monte Carlo simulation with a GARCH(1,1) model for the deviation and a skewed student-t distribution for the marginals, as seen on the BMW and Tesla examples, can properly predict the risk of securities in the stock market over a period of time, only by checking the time series of the product. However, there are certain situations and events that cannot be extracted from the time series, which could affect the Risk of a product.

On these specific scenarios, Samsung had announced at the end of the first period the release of the Samsung Galaxy S2, the top selling smart-phone of 2011 with 40 millions of units sold increasing its market share from 17.7% to 22.0%.

Facebook, on the 17th of March of 2018 got involved in a controversy where a whistleblower revealed that British political consulting firm Cambridge Analytica harvested 50 million Facebook profiles and used personal information taken without authorization. Leading, one month later, Mark Zuckenberg (Facebook CEO) to testify in front of the senate about facebook's data privacy. Events that created two drops in the prices of the company at the stock market.

The two events were widely covered on the net by news agencies, electronic newspapers, technology web pages, social networks... Which could have been used, along with the regression models, as input data to predict the risk of the securities in a more realistic way.

Table 4 Value at Risk before and after the interval including interdependencies

VaR(%)	BMW	Tesla	Facebook
Before	−29.48	−53.81	−26.70
After	−37.66	−57.55	−30.55

Table 5 Value at risk of the portfolio before and after the interval

Portfolio VaR(%)	Single	Dependencies
Before	−36.14	−36.30
After	−41.31	−41.5

The unpleasant sentiments that Facebook users were feeling against the application was expressed in several social networks. A sentiment that could have been used to predict the increase of 13% of the VaR the next days. In the same way, the pleasant sentiment achieved by the good reviews of the Samsung galaxy S2 could have predicted the decrease of 23% of the electronics company shares.

When inter-dependencies with multivariate t-copulas come into account, it has been proven that non-correlated securities like Facebook, BMW or Tesla don't increase the overall Risk of the portfolio (slightly increased by the small correlation between the two car manufacturers). Which proves, that this kind of events like the Facebook controversy, are also not noticed until the market movement happens on portfolio level.

In addition, on 3rd January 2018, the regulation MiFID II came into force. Containing a group of laws and directives that provide harmonized regulation for investment services and consumers protection. The daily monitoring and reporting of investor's risk is one of the outcomes of this implementation. It prohibits financial entities to invest for their clients, with a different level of risk than the expected. A law, that conflicts with the impossibility of properly predicting the risk level in certain scenarios.

7 Conclusions, Limitations and Future Work

The result of this investigation is a development of an intelligent automated form to provide basically some prior knowledge of the likely risk scenarios that can face different finance agents. As exposed above, the IFM method can predict the risk of securities, if no recent drastic changes in the market are given, in order to influence in a positive way customer perceptions, preferences and final choices to become more rational and unbiased.

Some limitations and restrictions have been found when types of events such as drastic changes in the market occur. The model is insufficient concerning its effectiveness when companies subjected to the media pressure come across contentious issues.

A future research is proposed by the authors to make this risk calculation method even more accurate and overcome the constraints mentioned by creating a factor model that would combine a regression model (like the IFM model) and the indexes extracted from a sentiment analysis. With this new approach, not only the past of the financial product would be taken into account, but also the present, which it is assumed to be the most determining for drastic changes in the market. The wide usage of social networks is giving the opportunity everybody to comment, rate, criticize or appreciate any kind of event. Also, the stock exchange related events. In a more technical point of view, social networks can be seen as a gigantic database of society's opinion. The progress in natural language processing, text analysis, machine learning and computational linguistics has provided the last years a wide range of tools to analyze and extract sentiment from texts. Sentiments that can be modeled as quantitative data. In that case, the large amount of data stored in the social networks, specifically related to companies of the stock exchange which can be used to extract quantitative indexes that would reflect the sentiment of the society in front of it.

Acknowledgements The authors are pleased to acknowledge the support of the Spanish Ministry of Economy, Industry and Competitiveness (Grant id.: TURCOLAB ECO2017-88984-R). Also, this work was supported by the Spanish Ministry of Economy and Competitiveness under contract TIN2017-84553-C2-2-R. Finally, some of the authors are members of the research group 2017-SGR363, funded by the Generalitat de Catalunya.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

References

Alberga, D., Shalita, H., & Rami, Y. (2008). Estimating stock market volatility using asymmetric garch models. *Applied Financial Economics, 18,* 1201–1208.

Antonides, G., & Van der Sar, N. L. (1990). Individual expectations, risk perception and preferences in relation to investment decision making. *Journal of Economic Psychology, 11,* 227–245.

Bailey, W., Kumar, A., & Ng, D. (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics, 102,* 1–27.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: the common stock investment performance of individual investors. *The Journal of Finance, 55,* 773–806.

Barber, B. M., & Odean, T. (2001). Boys will be boys: gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics, 116,* 261–292.

Bernard, D., Cade, N. L., & Hodge, F. (2018). Investor behavior and the benefits of direct stock ownership. *Journal of Accounting Research*, *56*, 431–466.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2,* 1–8.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews, 11,* 143–172.

Demarta, S., & McNeil, A. J. (2005). The t Copula and Related Copulas. *International Statistical Review / Revue Internationale de Statistique*, *73*(1), 111–129.

Embrechts, P., Lindskog, F., & McNeil, A. (2001). Modelling dependence with copulas and applications to risk management. *Department of Mathematics, ETHZ.*

Georgarakos, D., & Fürth, S. (2015). Household repayment behavior: The role of social capital and institutional, political, and religious beliefs. *European Journal of Political Economy, 37,* 249–265.

Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. In *4th international AAAI conference on weblogs and social media*.

Glasserman, P. (2003). *Monte carlo methods in financial engineering*. Springer, New York.

Gueant, O. (2012) Computing the value at risk of a portfolio: Academic literature and practionners'response.

Halko, M.-L., Kaustia, M., & Alanko, E. (2012). The gender effect in risky asset holdings. *Journal of Economic Behavior & Organization, 83,* 66–81.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*(3), 705–730.

Heimer, R. Z. (2014). Friends do let friends buy stocks actively. *Journal of Economic Behavior & Organization, 107,* 527–540 (**Empirical Behavioral Finance.**).

Hoffmann, A. O. I., Post, T., & Pennings, J. M. E. (2013). How investor perceptions drive actual trading and risk-taking behavior. *Journal of Behavioral Finance*, *16*, 94–103.

Jacobsen, B., Lee, J. B., Marquering, W., & Zhang, C. Y. (2014). Gender differences in optimism and asset allocation. *Journal of Economic Behavior & Organization, 107,* 630–651 (**Empirical Behavioral Finance.**).

Joe, H., & Xu, J.J. (1996). The estimation method of inference functions for margins for multivariate models. *Department of Statistics, University of British Columbia*.

Jondeau, E., Poon, S. H., & Rockinger, M. (2007). *Financial modeling under non-gaussian distributions*. Springer Science & Business Media.

Jorion, P. (1997). *Value at risk: the new benchmark for controlling market risk*. McGraw-Hill. ISBN 9780786308484.

Kiymaz, H., Öztürkkal, B., & Akkemik, K. A. (2016). Behavioral biases of finance professionals: Turkish evidence. *Journal of Behavioral and Experimental Finance, 12,* 101–111.

Knutson, B., Wimmer, E., Kuhnen, C., & Winkielman, P. (2008). Nucleus accumbens activation mediates the influence of reward cues on financial risk taking. *NeuroReport, 19,* 509–513.

Kuhnen, C., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron, 47,* 763–770.

Kuhnen, C., & Knutson, B. (2011). The influence of affect on beliefs, preferences, and financial decisions. *Journal of Financial and Quantitative Analysis, 46,* 605–626.

Kumar, S., & Goyal, N. (2015). Behavioural biases in investment decision making - a systematic literature review. *Qualitative Research in Financial Markets, 7,* 88–108.

Lambert, P., & Laurent, S. (2001). Modelling financial time series using GARCH-type models with a skewed student distribution for the innovations. Stat Discussion Paper; 0125 (2001) 21 p.

Linsmeier, T.J., & Pearson, N.D. (1996). *Risk measurement: an introduction to value at risk*. Finance 9609004, University Library of Munich, Germany.

McInish, T. H., & Srivastava, R. K. (1984). The nature of individual investors' heterogeneous expectations. *Journal of Economic Psychology,5*, 251–263.

McNeil, A. J., Frey, R., & Paul, E. (2005). *Quantitative risk management*. Princeton University Press.

Mugerman, Y., Sade, O., & Shayo, M. (2014). Long term savings decisions: financial reform, peer effects and ethnicity. *Journal of Economic Behavior & Organization, 106,* 235–253.

Nelder, J., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313.

Nelsen, R.B. (2006). An introduction to copulas. *Springer Series in Statistics*.

Nicolosi, G., Peng, L., & Zhu, N. (2009). Do individual investors learn from their trading experience? *Journal of Financial Markets, 12,* 317–336.

Pantelidis, T., & Pittis, N. (2005). Forecasting volatility with a garch (1, 1) model: Some new analytical and monte carlo results. *Research Gate*.

Pompian, M. (2017). *Risk tolerance and behavioral finance*. Advisor Perspectives, Inc., Lexington

Scherer, M., & Mai, J.-F. (2014). Financial engineering with copulas explained. *Palgrave macmillan, ed.*.

Statman, M., Thorley, S., & Vorkink, K. (2006). Investor overconfidence and trading volume. *The Review of Financial Studies, 19,* 1531–1565.

Tekçe, B., Yılmaz, N., & Bildik, R. (2016). What factors affect behavioral biases? Evidence from turkish individual stock investors. *Research in International Business and Finance, 37,* 515–526.

van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics, 101,* 449–472.

Weisbenner, S., & Ivkovich, Z. (2003). Local does as local is: Information content of the geography of individual investors' common stock investments. *The Journal of Finance*.
Wiener, Z., & Benninga, S. (1998). *Value-at-risk (var)*. Mathematicsa in Education and Research, 7.
Xu, J.J. (1996). Statistical modelling and inference for multivariate and longitudinal discrete response data (T). Retrospective Theses and Dissertations, 1919–2007. University of British Columbia.
Zhang, Jin, & Ng, Wing Long. (2010). *Exact maximum likelihood estimation for copula models*. Centre for Computational Finance and Economic Agents.

Chapter 4

Global discussion of results

A collection of ICT tools based on big data technologies and social media are presented. This investigations have introduced the use of big data in sectors like mining and leather which its application its scarce. This opens up new opportunities for research and provides a valuable source of information.

In accordance with the literature, the use of social media to build a sentiment analysis tool allows to identify the stakeholders, gives the chance to satisfy them and the possibility to engage with them over a long time.

One of our contributions is the development of two databases, composed by 2.090.000 tweets related with the mining and leather sectors that will serve for future projects in this areas helping to understand which are the most spoken topics about leather's and mining's CSR words, which are the opinions about them and where the industries are shifting in terms of CSR. In addition, both frameworks provides a platform for further research and application to other sectors. However, to take advantage, CSR researchers and practitioners need to be familiar with computational data collection techniques such as APIs and programming algorithms. Although the project has third party dependencies, the projects are easy to maintain and update as they make use of API calls.

The results demonstrated in this work provide a new perspective for understanding that CSR debate in the Twitter sphere is growing and that the open discussion is mainly centered on the impact of mining industries on the land and on ethical concerns referring to working conditions and occupational safety. For public administrations and governments, our findings seem to indicate the need for more strict environmental regulations, so that the negative impacts on the environment and society get reduced.

Apart from that, in the course of this work we also discovered that leather companies face a challenging competitive environment and must adapt and realign strategies to continue responding positively to environmental and sustainable issues. This indicates a need to take action in favor of leather industry in order to decide what is good and wrong, which skin or hide counts and provide added value and meaning to the manufactured leather. Additional data is gathered through the use of questionnaires to address gaps that are apparent in our data. Results show that leather companies are familiar with CSR practices, but there is further scope for improvement in terms of formalization of CSR procedures and measurement systems. Its analysis also reveals that the management of CSR activities improves with the diffusion of knowledge on CSR practices.

The IFM method as a financial instrument was also presented. It was proved that can properly predict the risk of securities in the stock market over a period of time, only by checking the time series of the product. However, the model is insufficient concerning its effectiveness when companies subjected to the media pressure come across contentious issues. This can happen in some situations and thus requires the incorporation of sentiment analysis models. Emerging developments in communication technologies and the IT revolution in recent years have forced news agencies, electronic newspapers, technology web pages to cover on the net these type of events which could have been used, along with the regression models, as input data to predict the risk of the securities in a more realistic way.

Chapter 5

General conclusions and future directions

This chapter summarizes the research work on ICT tools applied to Business and social media that are presented in this thesis and highlights the main findings. It also outlines a number of future research directions.

5.1 Conclusions

On the one hand, this thesis implemented different tools to analyze stakeholders' perceptions from the CSR perspective in two controversial industry sectors involving big data and AI technologies.

From the mining industry perspective, results supported that companies use Twitter to inform about different initiatives conducted to be more sustainable and environmentally responsible. Social actions seemed to be the second most discussed issue and the economic aspect of CSR the least commented topic in the Twitter sphere. Results also revealed a slight tendency to positive tweets (48%) in front of negative tweets (25%) and neutral tweets (28%). It was not possible to draw any strong conclusions that indicated whether environmental issues, social issues, or economic issues were seen as more positive and which ones are seen as more negative. Closer examination of the results revealed that the mining sector should improve the CSR disclosure and adopt a stakeholder engagement strategies based on social media due to the percentage of tweets posted by anonymous personal accounts accounted for 87% while those written by official accounts only reached 13%.

With regards to leather sector, around 74% of the companies surveyed were familiar with CSR management systems and many of them have undertaken some actions, particularly in relation with the environment. Most companies realized the benefits of having a management systems while others admitted unawareness or difficulties to implement them. However, the number of companies that adopted CSR procedures and measurement systems was relatively low, and the diffusion of sustainable leather standards was fairly sparse. These results are in line with a clear bias to negative sentiment tweets (around 43%) in front of mixed tweets (20%) and positive tweets account approximately just for 8%. *Sustainability* and *transparency* were less well received by the public while *green* and *human rights* were more supportive. Tweets posted by official users only reach 2,5% which in the same way than mining industry, need to be improved to involve stakeholder engagement.

Moreover, we found that leather industry is strongly linked with fashion, probably due to industries have turned leather into an all-season product. The debate is focused on the impact of leather industries against the planet and the economy and many stakeholders are arguing for greater environmental responsibility since they consider vegan alternatives such as organic biomaterials made from anything but animals.

On the other hand, a development of an intelligent automated tool to provide basically some prior knowledge of the likely risk scenarios that can face different finance agents is developed. It was demonstrated that the IFM method can predict the risk of securities, if no recent drastic changes in the market are given, in order to influence in a positive way customer perceptions, preferences and final choices to become more rational and unbiased. Nevertheless, the findings suggested that this does not represent a viable mechanism when drastic changes in the market occur.

5.2 Future research directions

These methods were limited to twitter but could be extended for other social networks and thus compare results between different social networks and to conduct more detailed study of the CSR mining and leather debate. They could further be refined by also considering the possibility to add more languages and more filters. Our research has some practical limitations, since it only considered the Tweets collected within a specific period of time and the Tweets written in English and Spanish. In addition, we developed a predefined list of specific initial parameters to filter the Tweets. In this sense it would be interesting for further research to include more languages and more filters in order to extend the list of initial words to search within the Tweets. It would be also interesting to conduct similar studies analysing correlations between topics and words, to correlate topics and profiles of users (professional and non–professional accounts, geographical origin, ...).

In future work, it may be useful to study particular aspects of a potential problem that real-time sentiment monitoring applications may imply, which could be the tendency for managers to view a continuing stream of analysis and reports without making any decisions or taking any action. "Sentiment is up . . . no, it's down . . . alright, it's back up again!" For ongoing monitoring work, there should be processes for determining when specific decisions and actions are necessary – when, for example, data values fall outside certain limits. Such information helps to determine decision stakeholders, decision processes and the criteria and time frames for which decisions need to be made.

The full potential of the IFM method had not been proven, and hence a future research is proposed to make this risk calculation method even more accurate and overcome the constraints mentioned by creating a factor model that would combine a regression model (Like de IFM method) and the indexes extracted from a sentiment analysis. It could be formulated as follows:

$$VaR = VaR_{ifm} * \alpha + VaR_{sentiment} * \beta$$

With this new approach, it would be possible to take into account the current status of a financial product and calculate the risk of an equity according to the sentiment of the society in front of it.

Bibliography

Ali, Imran, Ana Isabel Jiménez-Zarco, and Marta Bicho (2015). "Using social media for CSR communication and engaging stakeholders". In: *Corporate social responsibility in the digital age*. Emerald Group Publishing Limited.

Antolín, Mariano Nieto and Roberto Fernández Gago (2004). "Responsabilidad social corporativa: la última innovación en managementResponsabilidad social corporativa: la última innovación en management". In: *Universia Business Review* 1, pp. 28–39.

Bae, Kee-Hong, Sadok El Ghoul, and Omrane Guedhami Zhaoran (Jason) Gong (2021). "Does CSR matter in times of crisis? Evidence from the COVID-19 pandemic". In: *Journal of Corporate Finance* 67, p. 101876. ISSN: 0929-1199. DOI: `https://doi.org/10.1016/j.jcorpfin.2020.101876`. URL: `https://www.sciencedirect.com/science/article/pii/S0929119920303205`.

Barber, Brad M and Terrance Odean (2001). "Boys will be boys: Gender, overconfidence, and common stock investment". In: *The quarterly journal of economics* 116.1, pp. 261–292.

Beaver, Doug et al. (2010). "Finding a needle in haystack: Facebook's photo storage". In: *9th USENIX Symposium on Operating Systems Design and Implementation (OSDI 10)*.

Benthaus, Janek, Marten Risius, and Roman Beck (2016). "Social media management strategies for organizational impression management and their effect on public perception". In: *The Journal of Strategic Information Systems* 25.2, pp. 127–139.

Bestratén, M and L Pujol (2003). "Responsabilidad social de las empresas (I): conceptos generales". In: *Instituto Nacional de Seguridad e Higiene en el Trabajo. NTP* 643, p. 2003.

Bortree, Denise Sevick and Trent Seltzer (2009). “Dialogic strategies and outcomes: An analysis of environmental advocacy groups’ Facebook profiles”. In: *Public relations review* 35.3, pp. 317–319.

Brown, Brad, Michael Chui, and James Manyika (2011). “Are you ready for the era of ‘big data’”. In: *McKinsey Quarterly* 4.1, pp. 24–35.

Carroll, Archie B (1999). “Corporate social responsibility: Evolution of a definitional construct”. In: *Business & society* 38.3, pp. 268–295.

Commission, European (Mar. 2006). *Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee - Implementing the partnership for growth and jobs : making Europe a pole of excellence on corporate social responsibility.* URL: `https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:52006DC0136`.

Davenport, Thomas H (2014). “How strategists use “big data” to support internal business decisions, discovery and production”. In: *Strategy & Leadership.*

Du, Shuili and Edward T Vieira (2012). “Striving for legitimacy through corporate social responsibility: Insights from oil companies”. In: *Journal of business ethics* 110.4, pp. 413–427.

Galán, JI (2008). “Responsabilidad social corporativa, cambio institucional y gobierno: introducción y panorámica”. In: *Revista Europea de Dirección y Economı a de la Empresa* 17.3, p. 7.

Iazzi, Antonio et al. (2020). “Communicating the stakeholder engagement process: A cross-country analysis in the tourism sector”. In: *Corporate Social Responsibility and Environmental Management* 27.4, pp. 1642–1652.

ISO (2010). URL: `https://www.iso.org/obp/ui/#iso:std:iso:26000:ed-1:v1:en:sec:1`.

Kanagaraj, J et al. (2015). “Eco-friendly waste management strategies for greener environment towards sustainable development in leather industry: a comprehensive review”. In: *Journal of Cleaner Production* 89, pp. 1–17.

Kelleher, Tom (2006). *Public relations online: Lasting concepts for changing media.* Sage.

Kiymaz, Halil, Belma Öztürkkal, and K Ali Akkemik (2016). “Behavioral biases of finance professionals: Turkish evidence”. In: *Journal of Behavioral and Experimental Finance* 12, pp. 101–111.

Kühl, Niklas et al. (2019). “Literature vs. Twitter: Empirical insights on customer needs in e-mobility”. In: *Journal of cleaner production* 213, pp. 508–520.

Liang, Po-Wei and Bi-Ru Dai (2013). "Opinion mining on social media data". In: *2013 IEEE 14th international conference on mobile data management*. Vol. 2. IEEE, pp. 91–96.

Martínez, Carmen Valor and Marta de la Cuesta González (2003). "Responsabilidad social de la empresa. Concepto, medición y desarrollo en España". In: *Boletín Económico De ICE, Información Comercial Española* 2755, pp. 7–20.

McAfee, Andrew et al. (2012). "Big data: the management revolution". In: *Harvard business review* 90.10, pp. 60–68.

Nicolosi, Gina, Liang Peng, and Ning Zhu (2009). "Do individual investors learn from their trading experience?" In: *Journal of Financial Markets* 12.2, pp. 317–336.

Noland, James and Robert Phillips (2010). "Stakeholder engagement, discourse ethics and strategic management". In: *International Journal of Management Reviews* 12.1, pp. 39–49.

O'Riordan, Linda and Jenny Fairbrass (2014). "Managing CSR stakeholder engagement: A new conceptual framework". In: *Journal of business ethics* 125.1, pp. 121–145.

Perrey, J, D Spillecke, and A Umblijs (2013). "Smart analytics: How marketing drives short-term and long-term growth". In: *McKinsey Quarterly*, pp. 00425–3.

Rahman, Shafiqur et al. (2011). "Evaluation of definitions: ten dimensions of corporate social responsibility". In: *World review of business research* 1.1, pp. 166–176.

Redondo, H (2006). "El compromiso social de las empresas con las ciudades". In: *Deloitte mayo*.

Romera, Marı a Dolores Mellado (2006). *La gestión integrada de la calidad, el medio ambiente y la prevención de riesgos laborales en las OR*. EDITORIAL RAMÓN ARECES.

Steffen, Will et al. (2015). "The trajectory of the Anthropocene: the great acceleration". In: *The Anthropocene Review* 2.1, pp. 81–98.

Wamba, Samuel Fosso et al. (2015). "How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study". In: *International Journal of Production Economics* 165, pp. 234–246.

Xiang, Zheng and Ulrike Gretzel (2010). "Role of social media in online travel information search". In: *Tourism management* 31.2, pp. 179–188.

Xu, James Jianmeng (1996). “Statistical modelling and inference for multivariate and longitudinal discrete response data”. PhD thesis. University of British Columbia.