APPLICATION OF TEXT MINING TO THE ANALYSIS OF CLIMATE-RELATED DISCLOSURES

2020

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2035

Ángel Iván Moreno and Teresa Caminero

APPLICATION OF TEXT MINING TO THE ANALYSIS OF CLIMATE-RELATED DISCLOSURES

APPLICATION OF TEXT MINING TO THE ANALYSIS OF CLIMATE-RELATED DISCLOSURES [*]

Ángel Iván Moreno and Teresa Caminero

BANCO DE ESPAÑA

(*) The authors thank José Manuel Marqués, Sergio Gorjón, Ana Fernández, Alexandra Luccioni and Sadid Hasan for their comments and suggestions. The opinions and analyses are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Bank of Spain or the Eurosystem.

Documentos de Trabajo. N.º 2035

2020

ISSN: 1579-8666 (on line)

Abstract

In this article we apply text mining techniques to analyse the TCFD recommendations on climate-related disclosures of the 12 significant Spanish financial institutions using publicly available corporate reports from 2014 until 2019.

In our analysis, applying our domain knowledge, first we create a taxonomy of concepts present in disclosures associated with each of the four areas described in the TCFD recommendations. This taxonomy is then linked together by a set of rules in query form of selected concepts. The queries are crafted so that they identify the excerpts most likely to relate to each of the TCFD's 11 recommended disclosures. By applying these rules we estimate a TCFD compliance index for each of the four main areas for the period 2014-2019 using corporate reports in Spanish. We also describe some challenges in analysing climate-related disclosures. The index gives an overview of the evolution of the level of climate-related financial disclosures present in the corporate reports of the Spanish banking sector. The results indicate that the quantity of climate-related disclosures reported by the banking sector is growing each year. Besides, our study also suggests that some disclosures are only present in reports different than annual and ESG reports, such as Pillar 3 reports or reports on remuneration of directors.

Keywords: sustainability, sustainability data gaps, text mining, TCFD, Taxonomy and Ontology Management.

JEL classification: C81, G32, Q54.

Resumen

En este artículo aplicamos técnicas de minería de textos para analizar las recomendaciones del TCFD sobre la divulgación financiera relacionada con el clima de las 12 entidades significativas españolas, usando los informes corporativos disponibles públicamente desde 2014 hasta 2019.

En el análisis, aplicando nuestro conocimiento del área, creamos primero una taxonomía de conceptos presentes en la información reportada asociada a cada una de las cuatro áreas descritas en las recomendaciones del TCFD. Esta taxonomía se relaciona entre sí mediante un conjunto de reglas en forma de consultas que seleccionan conceptos. Las consultas se crean de manera que identifican los fragmentos que con mayor probabilidad están asociados a cada uno de los 11 aspectos recomendados para divulgar. Aplicando estas reglas, estimamos el índice de cumplimiento para cada una de las cuatro áreas principales para el período 2014-2019, usando informes corporativos en español. También describimos los retos que se presentan al analizar la divulgación financiera relacionada con el clima. El índice da una visión de la evolución del nivel de información relacionada con el clima presente en los informes corporativos del sector bancario español. Los resultados indican que la cantidad de información relacionada con el clima divulgada por los bancos crece cada año. Además, nuestro estudio también sugiere que hay información que solo está presente en informes diferentes de los informes anuales o de los informes ESG, como pueden ser los informes de Pilar 3 o los informes anuales de remuneraciones de consejeros.

Palabras clave: sostenibilidad, minería de textos, procesado de lenguaje natural, TCFD, cambio climático, informes corporativos, gestión de taxonomías y ontologías.

Códigos JEL: C81, G32, Q54.

1. Introduction

The European Directive 2014/95/UE, also called the non-financial reporting directive (NFRD), represents an important milestone towards improving the current corporate reporting to include not only the tangible assets but also the intangible assets, such as the communication, the culture the brand and the reputation, which require non-financial indicators, besides the traditional financial indicators (Instituto de auditores internos de España 2018). Recent studies indicate that the total value of the intangible assets can be greater than the value of the tangible assets (Brand Finance Institute 2017). Climate-related disclosures is one type of non-financial disclosures that investors have been pressing companies to include in their reports. Although initially this might have been mainly driven by an ecological activism caused by growing awareness of the impact of financial investments on the deterioration of the Earth, it also soon became clear that climate change was a source of risks as well as opportunities. (FIR -Forum pour l'investissement responsible 2016).

This interest on the financial impact of climate change prompted the G20 Finance Ministers and Central Bank Governors, in their Meeting in Washington in April 2015, to ask the Financial Stability Board (FSB) to "convene public- and private- sector participants to review how the financial sector can take account of climate-related issues" (G20 Finance Ministers and Central Bank Governors Meeting 16-17 April 2015). Underscoring this need, in September of the same year, Mark Carney, Chairman of the Financial Stability Board (FSB) at the time, performed a historic speech (Carney 2015) where he stressed the importance of company disclosures so that "better information – about the costs, opportunities and risks created by climate change – can promote timely responses" and introduced the Climate Disclosure Task Force (CDTF), later known as the TCFD (Task Force on Climate-related Financial Disclosures).

This speech acknowledged that Central Banks, in their mandate to protect financial stability, should consider the financial risk that climate factors create. It also made clear that climate-related corporate disclosures were key in assessing these risks.

In the financial sector in particular, in an analysis published in 2018, the Bank of England identified three broad categories in which banks were responding to climate-related risks: as being "responsible", focusing in the Corporate Social Responsibility (CSR) aspect to reduce reputational risk; as being "responsive", where climate change is viewed only from a short-term financial risk perspective and as being "strategic"; taking a long-term view of the financial risks involved (Bank of England 2018). Although the report focused on banks, these three categories could also be applicable to other sectors.

In Spain, corporate reporting is mainly regulated by the "Ley de Sociedades de Capital". Accordingly, most corporations (depending on their type, the requirements may vary) need to produce three main types of reports[1]: Annual Financial Statements, Corporate Governance Report and Management Report. The Management Report and the Annual Financial Statements are often put together in one single document. Listed companies are also required to create an Annual Report on Remuneration of Directors. As part of the Basel Framework[2] requirements, internationally active banks need to produce an additional prudential report following the Pillar 3 standard. In their Action plan for Sustainable Finance, the European Banking Authority (EBA) emphasized the need to disclose Environmental, Social and Governance (ESG) risks in this Pillar 3 report (EBA 2019). In particular, climate change is referred to as part of the environmental risk factors on which the EBA will especially focus in the first phase of the action plan.[3]

[1] Together with an Audit Report.

[2] The Basel Framework is the full set of standards of the Basel Committee on Banking Supervision (BCBS), which is the primary global standard setter for the prudential regulation of banks. The membership of the BCBS has agreed to fully implement these standards and apply them to the internationally active banks in their jurisdictions. (BIS 2020)

[3] Despite Climate Change being considered mainly an Environmental factor, both the EBA Action Plan and the European Commission Action Plan (European Commission 2018) highlight that "Environmental and social considerations are often intertwined, as especially climate change can exacerbate existing systems of inequality. The governance of public and private institutions […] plays a fundamental role in ensuring the inclusion of social and environmental considerations in the decision-making process."

The Annual Financial Statements are frequently produced in XBRL[4] format, as well as in pdf format, but they are often also part of a document normally called Annual Report which is meant for shareholders and other interested parties. This document is typically a colourful brochure that combines text, tables, pictures and charts and which also includes non-financial information. It sometimes follows a framework defined by the International Integrated Reporting Council (IIRC), which advocates for a single report where the financial and non-financial information is integrated in a cohesive way.

Before the transposition of the NFRD European Directive, there were no mandatory disclosures on many non-financial topics, including climate change, although many companies voluntarily created a separate ESG or CSR Report which served both as a marketing tool for investors and as a way to report non-financial information. While the mandatory reports are placed in the Commercial Register, the National Stock Market Commission (CNMV) or, in the case of the Pillar 3 report, at the Banco de España, there is no obligation to register the ESG reports in a centralized way. The NFRD Directive allows for publication of these reports in the respective corporate web sites. This actually means that the best way to find the main company reports that potentially contain climate-related disclosures is retrieving them from each individual corporate site.

There are several Spanish regulations related to the NFRD Directive, but for the purpose of this study it is worth mentioning three:

- Ministerial Order ESS/1554/2016, which, since the NFRD Directive allows for a separate report for the non-financial disclosures, establishes a voluntary procedure that allows publishing the ESG reports in a centralized way. This publicly available database was eventually called "Memorias de Responsabilidad Social de las Empresas" (MEMRSE[5]) and constitutes the Spanish Sustainability Register

[4] XBRL is the open international standard for digital business reporting, based on the XML standard and managed by a global not for profit consortium, XBRL International.

[5] https://expinterweb.mitramiss.gob.es/memrse/entrada/listadoMemoriasPublicadas.action

- 11/2018 Act, with reference BOE-A-2018-17989, which is the main law that transposes the NFRD directive. This enforces a specific set of non-financial disclosure for certain corporations, being applicable starting in the 2018 reporting period.

- Draft bill of Climate Change and Energy Transition, according to which the Spanish Macroprudential Authority Financial Stability Council (AMCESFI, for *Autoridad Macroprudencial Consejo de Estabilidad Financiera*)[6] will have to evaluate every two years the climate change risks for the financial sector.

6

According to the Internal Auditors Institute (2018) the five main standards for non-financial reporting that Spanish Corporations follow are: Global Reporting Initiative (GRI); Progress reports of the United Nations Global Compact; CDP (formerly Carbon Disclosure Project) reports; Sustainability Accounting Standards Board (SASB) -although this is mainly used in the US as per recommendation of the Securities and Exchange Commission (SEC) - and the IIRC framework.

The *KPMG Survey of Corporate Responsibility Reporting 2017* identified that 77% of the top 100 Spanish Corporations use the GRI standard as a reference for their ESG reports. Besides, according to the same report, more and more companies are reporting following the IIRC framework.

Although the recommendations of the TCFD are meant to be considered as part of the financial reporting, in practice they seem to be included in the ESG reports (KPMG 2017 and Marqués Sevillano and Romo González 2018). According to a recent report on practices among financial firms (IIF 2019), the type of documents where companies publish climate-related financial disclosures includes annual reports or climate position papers; ESG or CSR reports; integrated

[6] The AMCESFI is an inter-agency collegiate body attached to the head of the Ministry of Economic Affairs and Digital Transformation and participated by high-ranking officials from the said Ministry and the three national authorities with prudential regulatory and supervisory responsibilities for the Spanish financial system: the Banco de España, the Comisión Nacional del Mercado de Valores (CNMV, National Securities Markets Commission) and the Dirección General de Seguros y Fondos de Pensiones (Ministry's Directorate General for Insurance and Pensions Funds). (Banco de España 2020)

reports or Global Reporting Initiative (GRI) documents; and standalone TCFD reports. It is worth noting that in a 2017 response to the TCFD Report Consultation (GRI 2017), GRI acknowledged that 8 out of the 11 recommended disclosures corresponded, at least in part, to disclosures already established in the GRI Standards, which would explain why companies might try to follow the TCFD recommendations in their sustainability reports. This overlap also means that companies following GRI in their disclosures were probably aligned with some of the TCFD recommendations even before 2017.

The multiplicity of standards and recommendations and the fact that there is still no XBRL taxonomy for ESG reporting, which would make this data computer-readable, means that analysts have to go through usually lengthy pdf documents to extract the key information they require for their analysis, whether it is for supervisory, credit assessment, investment or other purposes.

There are many ESG/CSR reports studies in the literature, such as the already mentioned *KPMG Survey of Corporate Responsibility Reporting 2017. S*ome focus on sustainability or climate, such as Blacksun's study on the FTSE 100 ("The Ecosystem of Authenticity") or the Carbon Disclosure Standard Board (CDSB) report "First Steps: Corporate climate & environmental disclosure under the EU Non-Financial Reporting Directive". There are also studies on environmental disclosures specific to Spanish companies (Echave and Bhati 2010) or GRI disclosures of Mediterranean countries (Tarquinio, Raucci and Benedetti 2018). But most of these studies are based on manual work of reviewing the reports. A computer-aided study was also performed in 2008 (Doran and Quinn 2008) where they basically looked for the terms "climate change", "global warming" and "greenhouse gas" in the SEC 10K filings. On top of that, they also performed a manual analysis. El-Haj, et al. (2019) in their analysis of computational linguistics (CL) applied to the study of financial discourses, suggest using a Natural Language Processing (NLP) technique called Named entity recognition (NER) which they define as "an information extraction task that isolates and then classifies named entities into predefined categories such as person names, locations and

organizations", identifying it as one of the tools to gain traction in mainstream accounting and finance research and which they believe "offer promising ways to enhance the study of meaning in financial discourse". The simplest NER method they describe is using "handcrafted lists" or "bag of words" to detect these entities. A list-based approach was followed by Kravet (Kravet 2013) to detect risks in 10-K filings using keywords such as "can", "cannot", "could", "may" or "might".

NLP has been broadly used in financial research and, to a lesser extent, in central bank research, with a special focus on unsupervised techniques and applications related to sentiment analysis, topic modelling and complexity analysis (Bholat, et al. 2015). Supervised machine learning techniques for NLP applied to finance are less found in the literature and most of the studies are focused on sentiment analysis. Despite the significant body of work around financial NLP applications, as A. Luccioni indicates in her finance section of the Climate Change AI paper (Rolnick, et al. 2019), the field of climate finance has been largely neglected within the scope of financial research. Luccioni also argues that machine learning techniques can play a central role to improve this field. Together with H. Palacios they proposed an idea for a study on climate disclosures using state-of-the-art NLP tools (Luccioni and Palacios 2019), although at the point of this writing this proposal has not been further developed. The growing importance of this type of disclosures, together with the inherent difficulty of identifying them due to their heterogeneity and dispersed characteristic, seems to make the actual task of gathering these disclosures worthy of some kind of automation.

In their 2018 and 2019 (TCFD 2018 and 2019) Status Reports, the TCFD also made use of supervised machine learning techniques to identify areas of the corporate reports potentially containing information related to each one of 11 recommended disclosures related to the four recommendations (see Figure 1 for a summary of the recommendations and recommended disclosures). The process required an initial labelling of passages from 150 companies to train a

statistical model. Once trained, given an excerpt, the model would assign it a likelihood of being aligned with a recommended disclosure. By carefully adjusting the threshold, they created an index that allowed monitoring the level of compliance with the recommendations. In their 2019 Status Report they evaluated reports using AI from 2016 to 2019, including 104 banks of different jurisdictions and sub-industries, without giving additional details regarding the type of AI approach they employed.

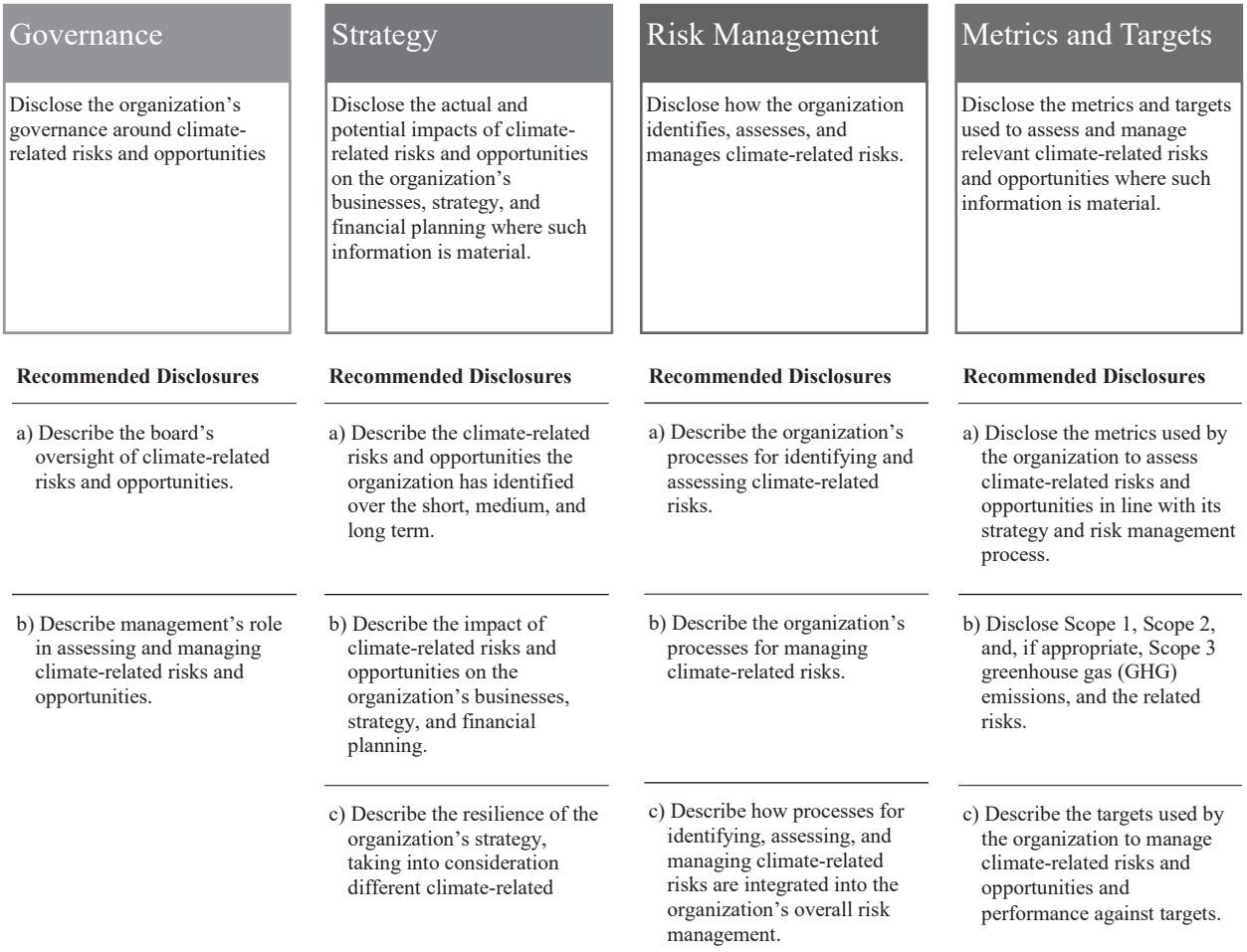

Figure 1: Recommendations and Supporting Recommended Disclosures (TCFD 2017)

The present study focuses on Spanish financial institutions, and in it we analyse 330 reports of the 12 significant Spanish institutions [7] to automatically estimate a TCFD compliance index. Instead

[7] As of 2020. Banks considered significant are under the ECB's direct supervision. To qualify as significant, banks must fulfil at least one of the criteria set out in the SSM Regulation and the SSM Framework Regulation regarding size, economic importance, cross-border activities or direct public financial assistance.

of using a statistical model, we use a rule-based model. The index is built based on search queries using key concepts to identify excerpts where the different recommendations are likely to be followed. These key concepts are part of a taxonomy initially created using the Spanish Sustainability Register. It was later adapted to fit some specificities of the banking sector reports.

To analyse the reports, we leverage NLP techniques using NER for information extraction. The NER method we use is based on a lexicon-based taxonomy that goes beyond the simplest "bag of words", taking into account lemmatization and regular expressions[8]. We then use the recognized entities in a second step following a rule-based approach to query the documents in order to create a compliance index based on the matches. The rule-based approach has the benefit of being easy to understand as well as being flexible in calibrating the model, since rules are easier to modify than a training set of labelled data. Besides, although we did not use a statistical model for the NER process, using a statistical model is still possible and it would allow us to have a hybrid model (a combination of a rule-based model and a statistical model). Finally the, automatic labelled data can easily be repurposed for helping the analysts identify specific topics within the domain, similarly to the 'human-in-the-loop' approach proposed by Luccioni and Palacios (2019). Specifically, when organizing related supervisory activities, this can contribute to a more efficient use of the available resources.

This approach is also similar to the one used for email surveillance for compliance purposes. In 2018 the Office of Compliance Inspections and Examinations (OCIE), which conducts the SEC's National Exam Program published a Risk Alert (OCIE 2018) where it highlighted some examples of practices that would assist in complying with the regulation. Within these examples was to have the ability to "compare postings to a lexicon of key words and phrases". Some examples of key words that are typically used are "can't talk", "political", "tip" or "in exchange" (NAIC 2018).

[8] Regular expressions (aka regexes) are a sequence of characters that define a search pattern. Regexes have far more capabilities than the usual wildcard characters and have a standard textual syntax for representing patterns for matching text.

The idea is that the presence of those key words in an email could be an indicator of employee misconduct. The lexicon can be organized into categories and additional rules might be used to have further information on the type of potential misconduct being identified. Although this approach can be prone to a significant number of false positives, it has the benefit of being easily interpretable, which is important in highly regulated environments.

The approach followed in this paper is also commonly used in enterprise search engines and is closely related to the technology referred by Gartner as "Enterprise Taxonomy and Ontology Management" (Gartner Inc 2016 and 2017). This paper also demonstrates that the current state of the art of the technology allows for research projects of intermediate size data using office-level personal computers.

In the following sections, first, we describe the selection process of the ESG reports available in the Spanish Sustainability Register. We used these reports to create the initial taxonomy, although we did not perform any additional analysis with them due to several inconsistencies. We then describe the selection process for the Bank reports, which we used to further enhance the taxonomy. The reports used in the analysis were not limited to Annual Reports and ESG reports, but were enriched with both national and banking-specific reports. These were as well the subject of our actual analysis of the financial institutions disclosures. Next, we define the rules to be used in the compliance index, also briefly describing the computer-aided process used to identify the text excerpts where the key areas of interest are located. Finally, we review some of the manual findings and challenges involved in analysing sustainability reports, and we evaluate the results of the calculated compliance index for the period 2014-2019.

2. Methodology

a. The Spanish Sustainability Register

Although any company is allowed to submit their ESG Reports to the Spanish Sustainability Register, we decided to select only those companies under the definition of "Sociedades de

Capital" which in Spain corresponds to "Sociedades de responsabilidad limitada", "Sociedades anónimas" and "Sociedades comanditarias por acciones" with more than 250 employees[9]. We did not exclude those companies that explicitly indicated that their ESG report did not contain environmental topics when they uploaded, since their ESG reports did not support that claim. This resulted in 118 reports of 53 companies for a period spanning from 2013 to 2018.

We found that the name of the reports can be misleading, since they can have names such as "Value Creation Report" or even "Annual Report" when they are actually ESG Reports. We can also find reports with the title "Progress Report" in reference to the United Nations Global Compact[10]. In any case, although Sustainability Reports typically use the GRI standards as a reference, there is no clear distinction between the content of ESG Reports, CSR Reports and Sustainability Reports besides the title given to them, which not always is fully aligned with the actual content. Therefore, we will refer to them globally as ESG Reports. In fact, in their latest review of their "Good Governance Code of Listed Companies", the CNMV has replaced the term CSR with ESG (CNMV 2020). Out of the companies reviewed, only a single one had a specific "Non-financial disclosures report". The remaining corporate reports that we used were either ESG Reports (37 companies) or Integrated Reports (15 companies).

The reports present in this Register were scarce, some companies did not upload them for most of the years, and even there were reports not matching the claimed description. This made this source of information unfit for a broader analysis, but it was still useful for the creation of an initial taxonomy. Since the process involved a manual review of specific sections of interest, it also uncovered some challenges and issues related to ESG reporting. The main ones are described later in the paper.

[9] This in line with the 11/2018 act, which, as mentioned in the introduction, is the main law that transposes the NFRD directive.

[10] The United Nations Global Compact is a call to companies everywhere to align their operations and strategies with ten universally accepted principles in the areas of human rights, labour, environment and anti-corruption, and to take action in support of UN goals and issues embodied in the Sustainable Development Goals. (UN Global Compact 2018)

b. Banks reports selection

After an initial taxonomy was created using the Sustainability Register as a baseline, this baseline was adapted with reports from the 12 significant Spanish financial institutions. In their 2018 Status Report, the TCFD evaluated the compliance with their recommendations by focusing on sustainability reports and financial filings of each company. It was further acknowledged that there was no single report where all disclosures could be found. They also included integrated reports, annual reports, "and other relevant documents" as needed, using reports of a previous year if the ones for 2017 were not available. In our analysis, to be able to compare disclosures coming from similar sources between financial institutions, only reports with a yearly frequency were selected. This means that documents such as policy reports and web-only content that was not annualized was not part of the study. This is in line with the TCFD recommendation that the disclosures should be issued "at least annually" (TCFD 2017). This also leaves out policy documents or other documents which are not updated annually and might be referenced in the main reporting document. Thus, the types of reports we used were:

1. Integrated Reports, Annual Reports or alternatively, Financial Statements, including the mandatory Management Report. These were categorized collectively as "Annual Reports".

2. ESG Reports. "ESG" is considered in the general sense, as mentioned in the previous section, which means that there were institutions that had more than one ESG Report for a given year.

3. Corporate Governance Reports

4. Reports on Remunerations of Directors

5. Pillar 3 reports

Note that not all of them are mandatory and the minimum number of reports available for a given year can be as low as three (Financial Statements, Pillar 3 Report and Corporate Governance

Report). In total 330 reports from the 12 significant banks were considered for the period between 2014-2019, having at least 3 reports for each institution and year.

These reports were manually retrieved from the corresponding corporate web sites. Whenever a mandatory report was not found in the corporate web site, it was retrieved from the CNMV web site. Once downloaded, the reports were classified, and they were processed in the same way as the ESG reports of the MEMRSE database. Besides, they were also analysed to adjust the taxonomy as described in the following section.

c. Taxonomy creation

To help with the creation of the taxonomy, we followed the workflow shown in Figure 2. First, we processed the pdf documents in order to extract their textual content. This was performed in a two-step process: first, using a commercial tool called *Kofax Power PDF Advanced* the pdf documents were turned into *MS Word* documents. Then, using a Python script, the *Word* documents were processed to extract the text and to partition it into excerpts. The intermediate *Word* step allowed for identification of paragraphs, tables and bullet points. The script treated each paragraph and bullet point as an excerpt. It also considered as an excerpt every table by itself, as well as any text content identified as a text box by the conversion tool. The script applied some additional heuristics to perform tasks such as trying to identify titles to consider them part of the next contiguous excerpt.

Within this study, we also developed a tool to be able to perform full text search (FTS) on the set of excerpts. The tool tokenized[11] and indexed all excerpts and allowed quickly finding all instances where a given word was used. The number of resulting excerpts for the total of the 330 corporate reports of the 12 significant institutions was close to 570.000.

[11] The process of tokenization is used in NLP to identify each individual lexical item present in a text, and it typically refers to demarcating the words within a text.

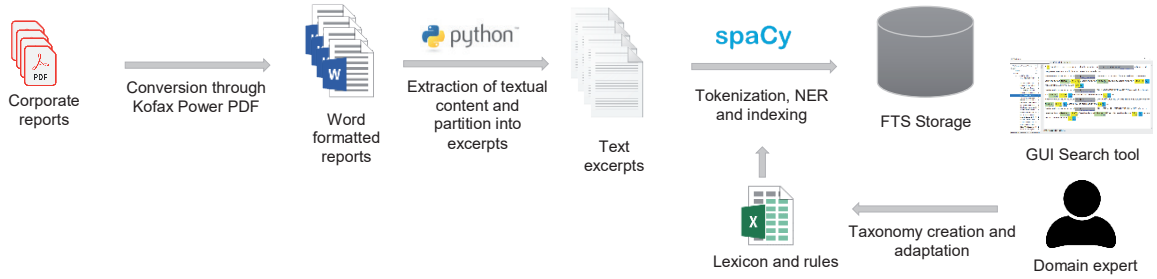

Figure 2: Representation of the workflow of our approach.

By using the FTS technique and manually reviewing the matching sections within the corporate reports, a taxonomy of concepts was created with the TCFD recommendations in mind. This way, for **Governance**, some of the specific concepts identified were: *board*[12], *management*[13], *sustainability committee*, *remuneration* and *periodicity*. For **Strategy**, some specific concepts were: *climate scenarios* and *temperature increment*. For **Risk Management**: *climate change risks*, *transition risks* and *physical risks*. For **Metrics and Targets**: *scope*, CO_2 *emissions* or CO_2 *units*. There were also other general concepts that could be used in any of the categories in combination with other concepts, such as: *climate change*, *sustainable* or *value*. The latter was used to identify numerical quantities, and it is an example of a concept where a regular expression pattern is more effective than a list Table 1 shows the main concepts used in the different recommendations.

Each of the concepts was then linked to a lexicon created from the actual reports also adding potential variants using our own domain knowledge. For example, for directors, the list contained, among others, the Spanish equivalent of Chief Executive Officer or general manager. Table 2 shows a sample of the actual Spanish content of three of the concepts of the taxonomy.

[12] Board of Directors or Board refers to a body of elected or appointed members who jointly oversee the activities of a company or organization (TCFD 2017). In Spain this is known as "Consejo de Administración"

[13] Management refers to those positions an organization views as executive or senior management positions and that are generally separate from the board (TCFD 2017). In Spain this is referred as "alta dirección" or simply "dirección".

Recommendation	Recommended Disclosure	Sample of Concepts used
Governance	a. Board Oversight	board, remuneration, periodicity, follow-up
	b. Management's Role	management, remuneration, periodicity, follow-up, sustainability committee
Strategy	a. Risks and Opportunities	climate change risks, opportunities, transition risks, physical risks, lending, mortgage, green loans, short, medium, long, extreme climate, cost reduction
	b. Impact on Organization	standards, strategy, impact, reputational risks, reporting standards, technology use, renewable energy
	c. Resilience of Strategy	climate scenarios, temperature increase
Risk Management	a. Risk ID & Assessment Processes	climate change risks, opportunities, transition risks, physical risks, processes, reporting guidelines, legal risk, reputational risk, financial risk, regulation, international agreements
	b. Risk Management Processes	risk response, materiality, carbon pricing, litigation, extreme climate, renewable energies, transition costs
	c. Integration into Overall Risk Mgmt	Integrated management, identification, risk management control system
Metrics and Targets	a. Climate-Related Metrics	reduction, CO_2 emissions, waste, energy consumption, water consumption, fuel consumption, renewable energy, value
	b. Scope 1,2,3 GHG Emissions	Scope, CO_2 emissions, CO_2 units, intensity
	c. Climate-Related Targets	target, reduction, CO_2 emissions, waste, energy consumption, water consumption, fuel consumption, renewable energy, value

Table 1: Main concepts used in each of the 11 recommendations

The purpose of the lexicon was to label the excerpts accordingly. To simplify the process, a given word could only have one label. If a given sequence of words was present in a concept and a subsequence of these was present in another concept, the longest sequence always prevailed. For example, in relation to the *CO2 emissions* concept, which included "emisiones de CO2" (Spanish for *CO2 emissions*). There was another concept called *emissions,* which among others, included the word *emisiones* (Spanish for *emissions*). If "emisiones de CO2" appeared in an excerpt, it

would never be labelled as simply *emissions*, since there is a concept (*CO2 emissions*) with a longer sequence of matching words.

CO2 emissions	Climate change risks	management
emisiones de co2	*riesgos de cambio climático*	*director ejecutivo*
niveles de carbono emitido	*riesgos del cambio climático*	*comité de dirección*
emisiones de carbono	*riesgos derivados del cambio climático*	*dirección general*

Table 2: Sample of the actual Spanish content of three of the concepts of the taxonomy

In order to reduce the number of errors due to ambiguities, a special *not applicable* category was created. This was deemed necessary as there were words that could be used within an applicable concept as well as within a concept not useful for the current analysis. Often including one or two of the surrounding words was enough to distinguish the meaning. For example, *climate* was initially part of the lexicon related to the *climate change* concept, but soon it became clear that in the reports there were more usages of *climate* in a different sense (*working climate*, *climate survey*, *organizational climate* or *political climate*). This resulted in only considering *climate* in combination with other words in order for it to be applicable to the *climate change* concept (e.g. *climate change* or *energy and climate*).

Additionally, often the actual text of the GRI requirements is also included in the ESG reports. To prevent identifying the actual text of a GRI reporting requirement as the text to comply with the requirement, verbatim sentences of the actual requirements were included as part of the *not applicable* concept. A similar situation happened with other mandatory reports that included the disclosure requirement as part of the report (e.g. Remuneration of Directors Report).

The lexicon was further lemmatized and lowercased, to consider this way minor variations of the terms. The excerpts were then automatically processed using Python and labelled according to the defined taxonomy.

d. Rules definition

To define a taxonomy of named entities for the NER model, we started from the 11 TCFD recommended disclosures. Since they are quite broad in their scope, we further divided those disclosures into 91 fine-grained disclosures using our own domain knowledge based on the examples and guidelines described by the TCFD (TCFD 2017). For each of these fine-grained disclosures, we defined a rule to determine if the disclosure was present in any of the corporate reports. The output of each rule was a value between 0 and 3. A zero value indicated that the disclosure was not found, while a value of three meant that there was a high probability that the disclosure was properly reported. Values 1 and 2 indicated a lower degree of certainty that the disclosure was either present or properly reported. For example, within the "Board Oversight" recommendation, one of the potential specific disclosures was information on how frequently there was a follow-up with the Board with the corresponding committee or similar on climate-related topics. The rule associated with this disclosure had three queries with varying level of precision in mind. If there was a match with the first query, it was evaluated as a "3", if it was the second query or third query that matched, it was evaluated as a "2" or "1" respectively. In the instance of no match, it resulted in a "0".

Recommend.	Recommended Disclosure	Specific Disclosure	Query	Value
Governance	a. Board Oversight	Follow-up frequency with the Board	**board** AND **periodicity** AND **sustainability committee**	3
			board AND **periodicity** AND **climate change**	2
			board AND **periodicity** AND (**sustainable** OR **esg**)	1

Table 3: Sample case of three rules with a progression of flexibility in their search concepts.

Table 3 shows this sample case where the three rules show a decreasing progression of specificity in their search concepts. Note that in the event of matching multiple queries for a given specific

disclosure, only the maximum value was taken into account. It was also perfectly possible that the same excerpt would account for two different specific disclosures. This was considered acceptable, so the corresponding score was assigned to the different disclosures according to the resulting value of the rule. It is important to highlight that the number of matches did not influence the result. The rules would assign the same score regardless if there was one or multiple excerpts that matched.

The 4-level scoring system allowed for a greater granularity than a binary approach (the disclosure is present or not present). Besides, since each recommended disclosure actually can refer to a diversity of specific disclosures, the rule approach gives the possibility to additionally create a weighting schema, which we did not use in this paper, to modulate the importance of specific disclosures.

The queries were defined using a very simple query language based on a combination of concepts from the taxonomy. The concepts could only be combined using the boolean operators AND, OR and NOT. Parentheses were allowed to group operators together.

After several iterations and reviews, eventually, 205 queries were defined, using a total of 81 different concepts. These queries were used to identify 91 different fine-grained disclosures, each one linked to one of the 11 recommended disclosures. Not all specific disclosures had 3 queries. Some of them had only 1 or 2. Table 4 shows the number of specific disclosures and number of queries for each of the recommended disclosures.

As an example of one of the ideas that was used to guide the levels is the differentiated concepts of *climate_change* and *sustainable*. The first one was related to the explicit mention of climate impact, while the second one was more related to the impact to the environment and natural resources. Besides, the word "sustainable" with the meaning of "able to be maintained" in phrases like "sustainable growth" or "sustainable in time" was considered part of the *not applicable*

Recommendation	Recommended Disclosure	# Specific Disclosures	# queries
Governance	a. Board Oversight	8	19
	b. Management's Role	13	34
Strategy	a. Risks and Opportunities	11	24
	b. Impact on Organization	16	24
	c. Resilience of Strategy	1	3
Risk Management	a. Risk ID & Assessment Processes	6	13
	b. Risk Management Processes	11	20
	c. Integration into Overall Risk Mgmt	3	7
Metrics and Targets	a. Climate-Related Metrics	7	18
	b. Scope 1,2,3 GHG Emissions	8	24
	c. Climate-Related Targets	7	19

Table 4: Number of specific disclosures and queries for each of the recommended disclosures.

category. Frequently, the rules were designed so that the relation of a disclosure with *climate change* was evaluated with a higher score than when it appeared only in relation with *sustainable*.

This approach is similar to but slightly more flexible than the one used by TCFD in their manual review where, for the Governance recommendation, they considered that "if a company described board or management responsibilities related to sustainability or ESG programs, but did not explicitly state that those programs included climate-related issues, the company's disclosure was not considered as aligned with the recommended disclosures". (TCFD 2018).

To calculate the index for each recommended disclosure, the average scoring of the specific disclosures was calculated and scaled to a number between 1 and 10. This value was then averaged across all the institutions to obtain the total score. The results can be seen in Figure 3 and are discussed later in the paper. Besides calculating the score, we also analysed the percentage of times each report type was used for scoring. Figure 4 shows a chart with this information. Whenever multiple reports matched the corresponding query, a precedence list was followed to create the

chart, so that only one document was accounted. This means that only the first matching document of the following list was used in the chart: Annual Reports, ESG Reports, Corporate Governance Reports, Remuneration of Directors Report and Pillar 3 Reports.

3. Evaluation of manual review findings

Although the ESG reports can follow a reporting standard (as mentioned before, in Spain this is commonly the GRI standard), they are typically not audited at the same level as the Financial Statements, if audited at all. The Non-Financial Disclosures (NFD) reports are often audited through what is known as "limited assurance engagement". This is the lower level of the two levels defined in the NIEA 3000 standard, which is the standard adopted by the Instituto de Censores Jurados de Cuentas de España (ICJCE 2019).[14]

Among the different problems we have found while manually reviewing the reports of the MEMRSE database, it is worth mentioning:

- The ESG reports too often contain marketing wording. Paraphrasing Richard Howitt[15]'s comment in the foreword of the 2019 Research Report (Alliance for Corporate Transparency 2020): *The major extractives company who sets its objective as "The wellbeing of our people, the community and the environment is considered in everything that we do," exposes 'warm words' rather than concrete targets, which are espoused in too much of today's reporting*. This means that queries used to identify specific disclosures needed to be carefully crafted to actually find the appropriate excerpts.

- It is not uncommon to find that a value for CO_2 emissions reported one year is different from the one that appears as reported for that year in the report of the following year. For,

[14] This lower level is targeted towards reducing the risk to an "acceptable level" to allow expressing the opinion of the auditor in a negative form. The higher level is targeted towards reducing the risk to a "reasonable level" to allow expressing the opinion of the auditor in a positive form.

[15] MEP with responsibility for parliamentary negotiation of the Non-Financial Reporting Directive in 2014. Chief Executive Officer, International Integrated Reporting Council (2016-2019)

example, a company may report 112 t of CO_{2e} in 2014. When looking at the report of 2015 the emissions of 2014 may appear as 113 t of CO_{2e}. Greenhouse Gas emissions (GHG) calculation is not an easy task, especially for big companies, and information required to perform the calculation might be delayed at the time of publishing the report, thus causing minor variations of the actual value. However, this change should be explained somehow in the report where the new value is reported. This would avoid confusion and would also allow stakeholders that actually keep track of those values to update their records accordingly. Unfortunately, more often than not, this clarification is not present. In any case, the objective of this semi-automated analysis was not to identify these errors, but whether the information was reported.

- As also identified by the TCFD (TCFD 2018), companies often do not describe the reasons for a given climate-related project, making it difficult for investors to understand the relevance of this project in relation to the company strategy. For example, when a company states that it "has multiplied its sustainability projects" without further information, leaving aside the fact that there is no information of which of these projects is related to climate-change, it is also not clear the benefit of these projects to the organization, with each project even potentially having a different strategic objective.

- There are instances where the climate related context is not fully contained in a given section, especially with risk management (TCFD 2018). Since sections are typically further divided into excerpts, this situation would prevent the technique used in this paper from identifying those disclosures.

On top of that, we have to consider some of the technical problems presented by typical ESG reports, among others:

- Since the reports are meant to be read by people, numerical information is often provided in the form of charts or infographics. This representation, although visually appealing and

sometimes easier to understand, is difficult to transform into a cohesive textual excerpt that could be further processed by a machine. Extraction and Optical character recognition (OCR)[16] tools might be able to identify the text pieces embedded in the images, but since the relationship between these text pieces is a spatial one, it will be lost when only considering the text itself, often losing the context of each individual text piece.

- Tables are also another common format to disclose numerical information. Similarly to infographics, companies tend to create visually appealing tables. The more imaginative and aesthetically designed a table is, the more challenging it is for a machine to identify it as a table. Sometimes, due to the distinct visual and spatial differentiation of the row or column headers in relation to the rest of the table, the Kofax conversion tool identified the headers and the body as independent elements and represented them in Word as two different tables, as a text-box and a table or even an image and a table. Although we added some hand-crafted rules to the processing script to cover some cases, this could cause the original table to end up divided into two excerpts, with the header in one and the body in another. Since each excerpt was treated independently by the rules, this also meant that the body of the table was deprived of the context implied in the headers.

- The pdf format is meant to be a description of a fixed-layout document as opposed to a reflowable format. Consequently, page breaks are clearly delimited. This often causes paragraphs to be broken even in mid-sentence. Although some heuristics were used to reduce this problem, there were always cases where a paragraph ended up divided into two excerpts. This problem was compounded whenever tables were involved since it was difficult to detect whether there were actually two tables or one continuous table across two pages.

[16] Optical Character Recognition is the process of converting images of text into machine-encoded text.

4. Evaluation of TCFD Compliance Index

The resulting value of the compliance index is shown in the chart of Figure 3. The bars represent the aggregated index for each *Recommended disclosure* while the lines show the range of values of the 12 institutions with marks signalling the index of each individual bank. The vertical areas highlight the limits of the 4-level scoring mechanism used in the analysis.

Figure 4 shows the distribution of report types where the disclosures were found. Only one report per disclosure is accounted for, following the priority indicated in the legend. The length of the bar represents the percentage of specific disclosures actually found in the reports for all banks, regardless the score obtained for each specific disclosure.

Governance Observations

a) Board Oversight

The rules used in this recommended disclosure were aimed to identify those excerpts with the presence of the *board* concept together with *climate change* related concepts, including concepts such as *periodicity*, *risks*, *remuneration* and *sustainability committee*. The chart in Figure 3 shows a progressive improvement of reporting from the apparently lowest level of 2015. Both 2014 and 205 have an aggregated score below the first scoring level, both in terms of number of specific disclosures and in terms of the score obtained by the defined rules. There seems to be more and more involvement of the respective Boards in climate-related issues. As can be seen in the chart of Figure 4, while Annual Reports and ESG Reports contain the largest amount of specific disclosures for other recommended disclosures, in this case, Corporate Governance, Pillar 3 and, increasingly, the Remuneration of Directors reports display a similar proportion of specific disclosures. As mentioned in a previous section, the fact that the rules use a flexible scoring in relation to the *Climate change* concept to accommodate the *Sustainable* concept, probably has the effect of showing a higher compliance level than what would be expected from the analysis performed by the TCFD in their Status reports.

b) Management Oversight

The rules used in this recommended disclosure were aimed to identify those excerpts with the presence of the *management* concept together with most of the concepts included in the previous disclosure with minor variations. The results displayed in Figure 3 do not seem to have a clear trend, probably because the Spanish reporting regulation is more focused on Governance disclosures at the Board level than at the Management levels. This is supported by the fact that the CNMV's *Code of good Governance* focuses on disclosures related to the Board (CNMV 2020). Overall management oversight ranks as the second lowest scored disclosure, with the aggregated score not surpassing the first level in any of the analysed years.

Strategy Observations

a) Risks and Opportunities

The rules used in this recommended disclosure were aimed to identify those excerpts with the presence of the general *risks* and *opportunities* terms in relation to *climate change* considering *short, medium* and *long* terms, including *transition risks* and *physical risk*, as well as specific risks and opportunities such as *extreme climate*, *cost reductions, lending* risks and *green products*. As with Board Oversight, the results in Figure 3 show a progressive improvement of reporting, but with a lowest located in 2014, also both in terms of number of specific disclosures and in terms of the score obtained by the defined rules. The lower margin also has an ascending trend, pushing the scores to the right side of the chart. In the chart of Figure 4, it is noteworthy the increased proportion of specific disclosures found in Pillar 3 reports, which should not come as a surprise since one of the main purposes of Pillar 3 reports is to provide stakeholders with information on the bank's material risks

b) Impact on Organization

The rules used in this recommended disclosure followed a variety of patterns. There were rules aimed to identify the presence of the *risks* concepts in relation to *climate change* and *impact,*

strategy or *objectives,* with a specific rule for *reputational risks*. There were also rules to find references to *climate change reporting standards.* Finally, there were rules to find specific impacts such as *technology use* or *renewable energy usage* in relation to *climate change*. The chart in Figure 3 shows a trend very similar to Board Oversight, with a lowest located also in 2015. As seen in Figure 4, Pillar 3 reports keep gaining importance in this area. As with the previous disclosure, the lower margin has an ascending trend, making the total score very close to the edge of the third lane. We interpret this as indicating that banks are increasingly reporting that climate-change issues are causing an impact in their organization such as adapting using new technologies or performing specific actions to reduce their reputational risk.

c) Resilience of Strategy

This recommended disclosure had only one rule, albeit with 3 queries, focused on identifying reporting of *climate scenarios* on *temperature increase*. The chart of Figure 3 shows that no institution reached level 3, and most were at level 0, although there is a clear progression towards levels 1 and 2. Note that level 2 indicates a lower certainty that the information was actually disclosed, not necessarily that the information was disclosed with lower quality, but in any case this makes this disclosure the lowest scored of all.

Risk Management Observations

a) Risk ID & Assessment Processes

The rules used in this recommended disclosure were aimed to identify those excerpts with the presence *risks* and *processes* in relation to *climate change.* Also references to *regulation,* specific risk-related *reporting frameworks* and *international agreements* together with *climate change.* As seen in the chart of Figure 3, this area shows a slow growth from an already moderate level, with the lowest scores also moving to the right. The chart of Figure 4 shows that Pillar 3 reports are gaining importance progressively in this area. Corporate Governance

reports seem to be identified as a source of this disclosure in the early years, but their overall weight decreases in the last years.

b) Risk Management Processes

As per the TCFD Annex (TCFD 2017), both this recommended disclosure and disclosure a) of the Strategy recommendation are referred to the same table of "Examples of Climate-Related Risks and Potential Financial Impacts". This means that disclosures in these two areas will very likely share similar concepts, being sometimes difficult to establish a clear differentiation between the two. Besides, there is also an overlap between this recommendation and the previous one. In fact, in their overall observations of their 2018 Status Report, the TCFD actually merged these two recommendations. In any case, the rules used in this recommended disclosure were aimed to identify those excerpts where concepts such as *risk_response* or *materiality* in relation to *climate change* were present. There were additional rules aimed to find references to *carbon pricing, litigation* and *transition costs,* always linked to *climate change.* As shown in the chart of Figure 3, there is an upward trend in this disclosure, with the lowest score also increasing its value, while the chart of Figure 4 shows that close to a third of excerpts are found in Pillar 3 reports.

c) Integration into Overall Risk Management

The rules used in this recommended disclosure were aimed to identify those excerpts with the presence of *climate change* in relation to the *integrated management* concept and *risks* or *risk control system.* This was the recommendation with the second lowest number of rules, being difficult to identify a wider variety of concepts that could fit into this recommended disclosure. Figure 3 shows that, as with most disclosures, we find an upward trend, this time also with a minimum in 2015. It is not until 2018, that the aggregated score goes above the first level. As per Figure 4, Pillar 3 reports seem to start including this recommended disclosure from 2017.

Metrics and Targets Observations

a) Climate-Related Metrics

The rules used in this recommended disclosure aimed to identify common metrics such as *renewable energy*, *emission reduction*, *waste*, *energy consumption*, *water consumption* or *fuel consumption*. Excerpts with actual *values* were evaluated higher, but it is difficult to determine whether the *values* were always actually related to the metrics. Figure 3 shows that this is the recommended disclosure with the highest score overall.

b) Scope 1, 2, 3 GHG Emissions

The rules used in this recommended disclosure tried to differentiate between the 3 different *scopes* and between raw *emissions* and *intensity* of *emissions*, also giving more weight to the excerpts where actual *values* were reported. Due to the limited breadth of the rules, there were actually banks that had the highest possible score in this disclosure. Figure 3 shows that the trend from 2014 to 2019 is also upward, with the exception of 2017. The first level is only surpassed in 2016, 2018 and 2019. As seen in Figure 4, this disclosure appears mainly in excerpts of Annual Reports and ESG Reports.

c) Climate-related Targets

The rules used in this recommended disclosure are very similar to the ones in the Climate-Related Metrics disclosure, but with the addition of the *target* concept. Figure 3 shows that, while the trend of the values is mostly kept, the actual values suffer a steep drop compared to the previous recommended disclosure because of this, which seem to indicate that institutions are much keener to report on the climate-related metrics than on the targets. The drop seems to be more intense for 2014 and 2015. Only in 2018 and 2019 they are slightly above the first level, making this disclosure the third lowest scored. Pillar 3 reports have a residual relevance in this disclosure, while for the metrics disclosure their weight was significant.

Figure 3: Estimated Compliance Index for the recommended disclosures. The bars represent the aggregated index while the lines show the range of values of the 12 institutions with marks signalling the index of each individual bank. The vertical areas highlight the limits of the 4-level scoring mechanism used in the analysis.

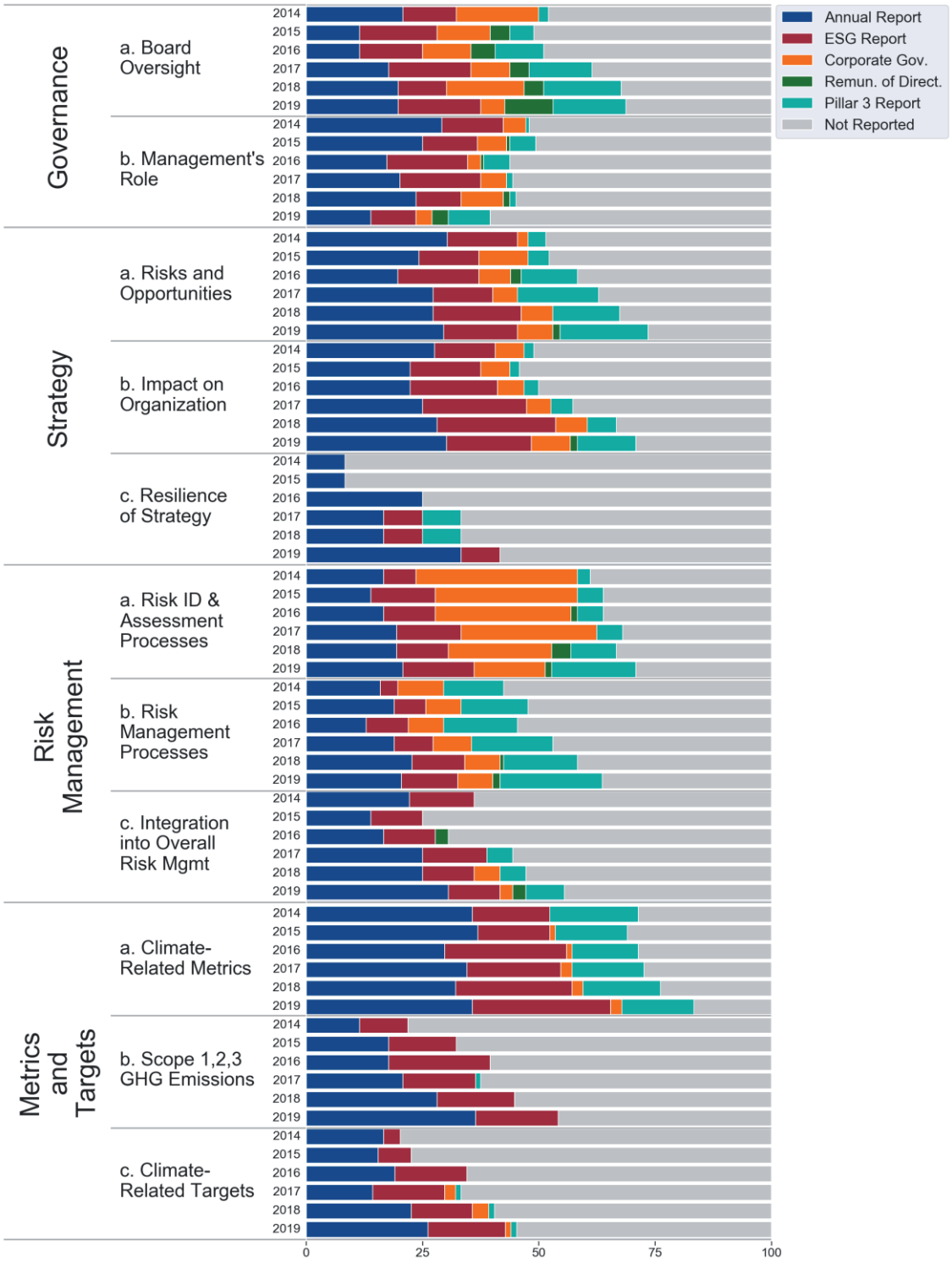

Figure 4: Distribution of report types where the disclosures were found. Only one report per disclosure is accounted for, following the priority indicated in the legend. The length of the bar represents the percentage of specific disclosures actually found in the reports for all banks, regardless the score obtained for each specific disclosure

5. Conclusions

In this paper, we used a rule-based NER approach to estimate an index that measures the level of compliance of the climate-related financial disclosures with the TCFD recommendations. We have applied this approach to estimate this TCFD compliance index analysing 330 reports of the 12 significant institutions of Spain using an NLP approach driven by NER.

Identifying the sections of the reports addressing specific climate-related financial disclosures can be seen as a text classification task, with really fine-grained categories. Besides, the number of excerpts that actually relate to a climate-related disclosure within the financial reports is sparse. Finally, some disclosures are much more present than others thus making the categories imbalanced as well. The fine granularity of the categories together with the scarcity and data imbalance present serious difficulties to an automatic text classification algorithm. Adding the fact that the language is not English but Spanish, also limit the availability of resources for a supervised machine learning approach. The approach presented in this paper relies on domain expertise and builds upon the traditional bag-of-words technique to create a rule-based NER model that not only helps analysts identify where potential disclosures are present, but also can be used for text classification purposes. When applied to the identification of climate related disclosures of the 12 significant Spanish banks we showed that banks are progressively improving their climate-related reporting, with three areas where banks seem to be lagging: Management Roles disclosures, Resilience of Strategy disclosures and Climate-Related Targets disclosures. It also showed that Annual Reports and ESG Reports should not solely be considered when evaluating the level of compliance with the TCFD recommendations, since national and sector regulations might indicate the need to include additional corporate reports to be able to have a better picture.

Due to the fact that the reporting documents are typically in pdf format and make extensive use of infographics, charts and, sometimes, graphically edited tables, the extraction of the textual information presents several challenges. This, together with the ambiguity when disclosing certain

information and other findings perhaps related to disclosure regulations not being as specific and strict as with financial information, make the task even more challenging.

The use of lexicons and rule-based approaches to text categorization is frequently discarded in favour of machine learning based techniques, although lexicons are still often used in sentiment analysis. In this paper, we have also shown a practical application of a rule-based approach that despite having the traditional problem of the need of domain expertise, it also tries to reduce the complexity problem through the creation of a taxonomy, providing this way higher flexibility and better interpretability. The rule-based model of the NER phase, could be enhanced with a statistical model, resulting in a hybrid model that could reduce the problem of words that have different meanings depending on the context as well as help with the scalability problem inherent to rule-based models. While the performance of a statistical NER model can typically be improved with additional training using annotated data, the performance of a rule-based model requires careful consideration of the existing rules as well as domain expertise. Is it also important to note that, in the case of a statistical model, domain expertise is required as well to be able to properly annotate training data, which can also be time-consuming. Besides, in instances where there is not enough data to train a statistical model, the effort might substantially increase in order to prevent the model from performing poorly, as synthetic data might need to be generated, with contexts where a given set of words is meant to be identified as a certain entity together with contexts where the same set of words is not. Alternatively, the identification of words to be added to the lexicon could also be aided with the application of word2vec techniques to obtain lists of similar words from which the domain expert can choose. This way, the approach described in the paper allows for a progressive improvement as analysts identify shortcomings.

The TCFD compliance index can be further enhanced with additional rules and a weighting schema of the different specific disclosures. Since the TCFD compliance index outlined in this paper has been crafted to identify specific fine-grained disclosures, equally weighted, if a different

weighting schema is used or if the index gets enhanced with additional disclosures, results will obviously vary and certain banks might perform differently in comparison, due to the human bias introduced when creating the rules.

The publication of the TCFD recommendations represent an important milestone towards standardization of the climate-related disclosures that are material to organizations. Following Carney's speech, by making sure that organizations provide better information about the costs, opportunities and risks created by climate change, timely responses can also be identified. But as long as climate-related disclosures do not reach the standardization level of financial statements, ideally in a machine readable format, analysts will have to carefully manually review the multiple reports published by corporations. The application of NLP techniques can greatly facilitate this task. Since there are specific corporate reports that are only in the Spanish language, it is important to develop NLP resources that help the advancement of the Spanish NLP applications. The approach presented in this paper can also be used to help building a baseline of training data for a machine-learning model that could overcome the scalability problem

References

Alliance for Corporate Transparency. “2019 Research Report.” 2020.

Banco de España. *Banco de España - Financial stability and macroprudential policy - AMCESFI.* 2020. https://www.bde.es/bde/en/areas/estabilidad/amcesfi/Autoridad_Macro_041deea38293961.html (accessed 7 15, 2020).

Bank of England. *Transition in thinking: The impact of climate change on the UK banking sector.* London: Prudential Regulation Authority, 2018.

Bholat, D, S Hansen, P Santos, and C Schonhardt-Bailey. *Text mining for central banks.* Bank of England, 2015.

BIS. *Basel Framework - bis.org.* 2020. https://www.bis.org/basel_framework/index.htm (accessed 7 15, 2020).

Brand Finance Institute. “Global Intangible Finance Tracker 2017.” 2017.

Carney, Mark. *Breaking the Tragedy of the Horizon – climate change and financial stability.* Bank of England, 2015.

CNMV. “Código de buen gobierno de las sociedades cotizadas.” 2020.

Doran, Kevin L., and Elias L. Quinn. “Climate Change Risk Disclosure: A Sector by Sector Analysis of SEC 10-K Filings from 1995-2008.” *34 N.C. J. Int'l L. & Com. Reg. 721*, 2008.

EBA. “EBA ACTION PLAN ON SUSTAINABLE FINANCE.” 2019.

Echave, Jon Otegui, and Shyam S. Bhati. “Determinants of social and environmental disclosures by Spanish Companies.” *GSMI Third Annual International Business Conference.* Michigan, USA, 2010. 55-68.

El-Haj, Mahmoud, Paul Rayson, Martin Walker, Steven Young, and Vasiliki Simaki. “In Search of Meaning: Lessons, Resources and Next Steps for Computational Analysis of Financial Discourse.” *Forthcoming in Journal of Business Finance and Accounting*, 2019.

European Commission. “Communication From The Commission To The European Parliament, The European Council, The Council, The European Central Bank, The European Economic And Social Committee And The Committee Of The Regions. Action Plan: Financing Sustainable Growth.” Brussels, 2018.

FIR -Forum pour l'investissement responsible. “«Article 173-VI: Understanding the French regulation on investor climate reporting».” *FIR Handbook 1, Octubre.* 2016.

G20 CFSG. “Report to the Finance Ministers.” 2015.

G20 Finance Ministers and Central Bank Governors Meeting. “Communiqué.” Washington, 16-17 April 2015.

Gartner Inc. “Hype Cycle for Emerging Technologies, 2016.” 2016.

Gartner Inc. “Hype Cycle for Emerging Technologies, 2017.” 2017.

GRI. “GRI’s response to the FSB TCFD Report Consultation.” *Global Reporting Initiative.* 2 2017. https://www.globalreporting.org/standards/media/1379/item-10-submission-gri-tcfd-publication.pdf (accessed 06 09, 2020).

Harrison, Patrick. “Spacy IRL: Creating a next-generation financial dataset from scratch with NLP & active learning.” 2019. https://www.youtube.com/watch?v=rdmaR4WRYEM (accessed April 24, 2020).

ICJCE. “Guía de actuación sobre encargos de verificación del Estado de Información No Financiera.” 2019.

IIF. *Climate-related Financial Disclosures: Examples of Leading Practices in TCFD Reporting by Financial Firms.* The Institute of International Finance, 2019.

Instituto de auditores internos de España. “Auditoría Interna y la información no financiera.” Madrid, 2018.

KPMG. “KPMG Survey of Corporate Responsibility Reporting 2017.” 2017.

Kravet, Todd. “Textual Risk Disclosures and Investors’ Risk Perceptions.” *Review of Accounting Studies 18*, 2013: 1088-1122.

Luccioni, Alexandra, and Hector Palacios. “Using Natural Language Processing to Analyze Financial Climate Disclosures.” *Proceedings of the 36 th International Conference on Machine Learning.* Long Beach, California, 2019.

Marqués Sevillano, José Manuel, and Luna Romo González. “El riesgo de cambio climático en los mercados y las entidades.” *REVISTA DE ESTABILIDAD FINANCIERA, NÚM. 34*, 2018.

NAIC. “Best Practices Guide for Email Surveillance.” *National Association of Investment Companies.* 2018. https://naicpe.com/best-practices-guide-for-email-surveillance/ (accessed June 15, 2020).

OCIE. “OCIE Risk Alert - Electronic Messaging.” *SEC.gov | Observations from Investment Adviser Examinations Relating to Electronic Messaging.* 14 December 2018. https://www.sec.gov/files/OCIE%20Risk%20Alert%20-%20Electronic%20Messaging.pdf (accessed June 15, 2020).

OECD. “G20/OECD CHECKLIST ON LONG-TERM INVESTMENT FINANCING STRATEGIES AND INSTITUTIONAL INVESTORS.” 2014.

Rolnick, David, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, and Alexandre Lacoste. “Tackling Climate Change with Machine Learning.” 2019.

Tarquinio, Lara, Domenico Raucci, and Roberto Benedetti. “An Investigation of Global Reporting Initiative Performance Indicators in Corporate Sustainability Reports: Greek, Italian and Spanish Evidence.” *Sustainability*, 2018.

TCFD. “2018 Status Report.” 2018.

TCFD. “2019 Status Report.” 2019.

TCFD. “Implementing the Recommendations of the Task-Force on Climate-related Financial Disclosures.” 2017.

TCFD. “Recommendations of the Task Force on Climate-related Financial Disclosures.” 2017.

UN Global Compact. “UN Global Compact Progress Report.” 2018.

Zillman, John W. *A History of Climate Activities.* 2009.

https://public.wmo.int/en/bulletin/history-climate-activities (accessed April 20, 2021).

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

1930 MICHAEL FUNKE, DANILO LEIVA-LEON and ANDREW TSANG: Mapping China's time-varying house price landscape.
1931 JORGE E. GALÁN and MATÍAS LAMAS: Beyond the LTV ratio: new macroprudential lessons from Spain.
1932 JACOPO TIMINI: Staying dry on Spanish wine: the rejection of the 1905 Spanish-Italian trade agreement.
1933 TERESA SASTRE and LAURA HERAS RECUERO: Domestic and foreign investment in advanced economies. The role of industry integration.
1934 DANILO LEIVA-LEON, JAIME MARTÍNEZ-MARTÍN and EVA ORTEGA: Exchange rate shocks and inflation comovement in the euro area.
1935 FEDERICO TAGLIATI: Child labor under cash and in-kind transfers: evidence from rural Mexico.
1936 ALBERTO FUERTES: External adjustment with a common currency: the case of the euro area.
1937 LAURA HERAS RECUERO and ROBERTO PASCUAL GONZÁLEZ: Economic growth, institutional quality and financial development in middle-income countries.
1938 SILVIA ALBRIZIO, SANGYUP CHOI, DAVIDE FURCERI and CHANSIK YOON: International Bank Lending Channel of Monetary Policy.
1939 MAR DELGADO-TÉLLEZ, ENRIQUE MORAL-BENITO and JAVIER J. PÉREZ: Outsourcing and public expenditure: an aggregate perspective with regional data.
1940 MYROSLAV PIDKUYKO: Heterogeneous spillovers of housing credit policy.
1941 LAURA ÁLVAREZ ROMÁN and MIGUEL GARCÍA-POSADA GÓMEZ: Modelling regional housing prices in Spain.
1942 STÉPHANE DÉES and ALESSANDRO GALESI: The Global Financial Cycle and US monetary policy in an interconnected world.
1943 ANDRÉS EROSA and BEATRIZ GONZÁLEZ: Taxation and the life cycle of firms.
1944 MARIO ALLOZA, JESÚS GONZALO and CARLOS SANZ: Dynamic effects of persistent shocks.
1945 PABLO DE ANDRÉS, RICARDO GIMENO and RUTH MATEOS DE CABO: The gender gap in bank credit access.
1946 IRMA ALONSO and LUIS MOLINA: The SHERLOC: an EWS-based index of vulnerability for emerging economies.
1947 GERGELY GANICS, BARBARA ROSSI and TATEVIK SEKHPOSYAN: From Fixed-event to Fixed-horizon Density Forecasts: Obtaining Measures of Multi-horizon Uncertainty from Survey Density Forecasts.
1948 GERGELY GANICS and FLORENS ODENDAHL: Bayesian VAR Forecasts, Survey Information and Structural Change in the Euro Area.
2001 JAVIER ANDRÉS, PABLO BURRIEL and WENYI SHEN: Debt sustainability and fiscal space in a heterogeneous Monetary Union: normal times vs the zero lower bound.
2002 JUAN S. MORA-SANGUINETTI and RICARDO PÉREZ-VALLS: ¿Cómo afecta la complejidad de la regulación a la demografía empresarial? Evidencia para España.
2003 ALEJANDRO BUESA, FRANCISCO JAVIER POBLACIÓN GARCÍA and JAVIER TARANCÓN: Measuring the procyclicality of impairment accounting regimes: a comparison between IFRS 9 and US GAAP.
2004 HENRIQUE S. BASSO and JUAN F. JIMENO: From secular stagnation to robocalypse? Implications of demographic and technological changes.
2005 LEONARDO GAMBACORTA, SERGIO MAYORDOMO and JOSÉ MARÍA SERENA: Dollar borrowing, firm-characteristics, and FX-hedged funding opportunities.
2006 IRMA ALONSO ÁLVAREZ, VIRGINIA DI NINO and FABRIZIO VENDITTI: Strategic interactions and price dynamics in the global oil market.
2007 JORGE E. GALÁN: The benefits are at the tail: uncovering the impact of macroprudential policy on growth-at-risk.
2008 SVEN BLANK, MATHIAS HOFFMANN and MORITZ A. ROTH: Foreign direct investment and the equity home bias puzzle.
2009 AYMAN EL DAHRAWY SÁNCHEZ-ALBORNOZ and JACOPO TIMINI: Trade agreements and Latin American trade (creation and diversion) and welfare.
2010 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: Eurozone prices: a tale of convergence and divergence.
2011 ÁNGEL IVÁN MORENO BERNAL and CARLOS GONZÁLEZ PEDRAZ: Sentiment analysis of the Spanish Financial Stability Report. (There is a Spanish version of this edition with the same number).
2012 MARIAM CAMARERO, MARÍA DOLORES GADEA-RIVAS, ANA GÓMEZ-LOSCOS and CECILIO TAMARIT: External imbalances and recoveries.

2013 JESÚS FERNÁNDEZ-VILLAVERDE, SAMUEL HURTADO and GALO NUÑO: Financial frictions and the wealth distribution.
2014 RODRIGO BARBONE GONZALEZ, DMITRY KHAMETSHIN, JOSÉ-LUIS PEYDRÓ and ANDREA POLO: Hedger of last resort: evidence from Brazilian FX interventions, local credit, and global financial cycles.
2015 DANILO LEIVA-LEON, GABRIEL PEREZ-QUIROS and EYNO ROTS: Real-time weakness of the global economy: a first assessment of the coronavirus crisis.
2016 JAVIER ANDRÉS, ÓSCAR ARCE, JESÚS FERNÁNDEZ-VILLAVERDE and SAMUEL HURTADO: Deciphering the macroeconomic effects of internal devaluations in a monetary union.
2017 FERNANDO LÓPEZ-VICENTE, JACOPO TIMINI and NICOLA CORTINOVIS: Do trade agreements with labor provisions matter for emerging and developing economies' exports?
2018 EDDIE GERBA and DANILO LEIVA-LEON: Macro-financial interactions in a changing world.
2019 JAIME MARTÍNEZ-MARTÍN and ELENA RUSTICELLI: Keeping track of global trade in real time.
2020 VICTORIA IVASHINA, LUC LAEVEN and ENRIQUE MORAL-BENITO: Loan types and the bank lending channel.
2021 SERGIO MAYORDOMO, NICOLA PAVANINI and EMANUELE TARANTINO: The impact of alternative forms of bank consolidation on credit supply and financial stability.
2022 ALEX ARMAND, PEDRO CARNEIRO, FEDERICO TAGLIATI and YIMING XIA: Can subsidized employment tackle long-term unemployment? Experimental evidence from North Macedonia.
2023 JACOPO TIMINI and FRANCESCA VIANI: A highway across the Atlantic? Trade and welfare effects of the EU-Mercosur agreement.
2024 CORINNA GHIRELLI, JAVIER J. PÉREZ and ALBERTO URTASUN: Economic policy uncertainty in Latin America: measurement using Spanish newspapers and economic spillovers.
2025 MAR DELGADO-TÉLLEZ, ESTHER GORDO, IVÁN KATARYNIUK and JAVIER J. PÉREZ: The decline in public investment: "social dominance" or too-rigid fiscal rules?
2026 ELVIRA PRADES-ILLANES and PATROCINIO TELLO-CASAS: Spanish regions in Global Value Chains: How important? How different?
2027 PABLO AGUILAR, CORINNA GHIRELLI, MATÍAS PACCE and ALBERTO URTASUN: Can news help measure economic sentiment? An application in COVID-19 times.
2028 EDUARDO GUTIÉRREZ, ENRIQUE MORAL-BENITO, DANIEL OTO-PERALÍAS and ROBERTO RAMOS: The spatial distribution of population in Spain: an anomaly in European perspective.
2029 PABLO BURRIEL, CRISTINA CHECHERITA-WESTPHAL, PASCAL JACQUINOT, MATTHIAS SCHÖN and NIKOLAI STÄHLER: Economic consequences of high public debt: evidence from three large scale DSGE models.
2030 BEATRIZ GONZÁLEZ: Macroeconomics, Firm Dynamics and IPOs.
2031 BRINDUSA ANGHEL, NÚRIA RODRÍGUEZ-PLANAS and ANNA SANZ-DE-GALDEANO: Gender Equality and the Math Gender Gap.
2032 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Machine learning in credit risk: measuring the dilemma between prediction and supervisory cost.
2033 PILAR GARCÍA-PEREA, AITOR LACUESTA and PAU ROLDAN-BLANCO: Raising Markups to Survive: Small Spanish Firms during the Great Recession.
2034 MÁXIMO CAMACHO, MATÍAS PACCE and GABRIEL PÉREZ-QUIRÓS: Spillover Effects in International Business Cycles.
2035 ÁNGEL IVÁN MORENO and TERESA CAMINERO: Application of text mining to the analysis of climate-related disclosures.

BANCO DE **ESPAÑA**
Eurosistema

Unidad de Servicios Generales I
Alcalá, 48 - 28014 Madrid
E-mail: publicaciones@bde.es
www.bde.es