DISCLOSURES OF BANKS' SUSTAINABILITY REPORTS, CLIMATE CHANGE AND CENTRAL BANKS: AN EMPIRICAL ANALYSIS WITH UNSTRUCTURED DATA

Dario Aversa *

* Department of Agricultural, Food, and Forest Sciences (SAAF), University of Palermo, Palermo, Italy;
SDA Bocconi School of Management, Bocconi University, Milan, Italy
Contact details: University of Palermo, Viale delle Scienze, 13 - Building 4, First Floor, 90128 Palermo, Italy

OPEN ACCESS

How to cite this paper: Aversa, D. (2024). Disclosures of banks' sustainability reports, climate change and central banks: An empirical analysis with unstructured data. *Risk Governance and Control: Financial Markets & Institutions, 14*(1), 76–102.
https://doi.org/10.22495/rgcv14i1p6

ISSN Online: 2077-4303
ISSN Print: 2077-429X

Received: 19.09.2023
Accepted: 19.03.2024

JEL Classification: G10, G20, G30, K32
DOI: 10.22495/rgcv14i1p6

Abstract

Climate change will impact the economy in the long term but also in the medium run potentially affecting financial stability and the whole economic system (Nyambuu & Semmler, 2023). Some studies suggested that climate change may hit financial actors even if there are few studies related to the role of banks' sustainability disclosure and central banks' mandates. This paper examines the sustainability reports disclosures of the banks listed on the FTSE Italia All-Share Italian Stock Exchange for the year 2021–2020. It applies five kinds of multivariate techniques on unstructured data using Iramuteq (www.iramuteq.org) and SAS Viya (www.sas.com). The article tries to assess how and whether banks are disclosing transition and physical risk, and how and whether they conducted scenario analysis through the lens of the Task Force on Climate-related Financial Disclosures (TCFD) framework. Even though banks provided environmental, social, and governance (ESG) disclosure, they paid scant attention to furnishing climate change information to market and central banks, and to prioritize their complying with international frameworks. The article addresses the main research gap in the literature review and focuses on the role of disclosures and central banks' mandates, sustainability data gap and solutions through machine learning, analysis of deep uncertainty for monetary policy, and the use of scenario analysis when it is impossible to apply deterministic or probabilistic approaches and text mining for central banks in the context of unstructured data.

Keywords: Climate Change, Central Banks, Disclosure, Text Analytics, Quantitative and Qualitative Research

Authors' individual contribution: The Author is responsible for all the contributions to the paper according to CRediT (Contributor Roles Taxonomy) standards.

Declaration of conflicting interests: The Author declares that there is no conflict of interest.

Acknowledgements: The Author is very grateful to Enzo Baglieri, Fabrizio Castellucci, Agostino Di Ciaccio, Fiorenza Deriu, and Stefano Fantacone.

1. INTRODUCTION

Climate change by the United Nations Framework Convention on Climate Change (UNFCCC)[1] was defined as a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere, a change of the state of climate that can be identified (e.g., by using statistical test) by changes in the mean and/or the variability of its properties and that persist for an extended period, typically decades or longer, climate change can be due to natural internal processes or external forcing such as modulations of the solar cycles, volcanic eruptions and persistent anthropogenic changes in the composition of atmosphere or in land use.

Climate change is a phenomenon that closely concerns science (Stern et al., 2022; Hwang et al., 2021, Ramani, 2020) and from the point of view of economics it produces uncertainty (Minenna & Aversa, 2019; U.S. Governing Publishing Office Washington, 2021; Barnett et al., 2022), deep uncertainty (McPhail et al., 2020; Aversa, 2023a, Minenna & Aversa, 2019) potentially affect financial institution and banks' balance sheet (TCFD, 2017a, 2019, 2020; Eccles & Krzus, 2018, Aversa & Cincinelli, 2019, Aversa, 2023b; Aversa et al., 2022) therefore the analysis over the mandatory disclosures (Moreno & Caminero, 2022; International Sustainability Standards Board [ISSB], 2022; The Bank of England, 2022) of sustainability report (Directive 2014/95/EU) become much more important for investors and central banks (TCFD, 2020a[2], 2022; Feridun & Güngör, 2020) to face the problem.

To fight against this complex, multifaceted, and nonlinear[3] phenomenon (Aversa et al., 2022), a treaty named The Paris Agreement was adopted by 196 parties at the UN Climate Change Conference[4] (COP 21) with the goal to hold "the increase in the global average temperature well below 2°C above pre-industrial levels" (United Nations Climate Change, n.d.) and pursue efforts "to limit the temperature increase to 1.5°C above pre-industrial levels"[5].

As climate-related financial risks clearly represent a significant threat to monetary and financial stability (Network for Greening the Financial System [NGFS], 2020; Official Monetary and Financial Institutions Forum [OMFIF], 2020) because may cause supply and demand shocks, the international community has reacted and among other actions, established the Task Force on Climate-related Financial Disclosures (TFCD) in 2015 by Financial Stability Board to cope with climate change.

Climate change risks can be transferred to banks, non-financial corporations, and the economy as a whole via transmission channels (Basel Committee on Banking Supervision [BCBS], 2021). Climate-related financial risks "could potentially impact the safety and soundness of individual financial institutions and have broader financial stability implications for the banking system" (BCBS, 2020, p. 1), banks and the banking system are exposed to climate change through macro-[6] and microeconomic[7] transmission channel[8] (BCBS, 2021).

In this regard, the TCFD (2017a) divides risks into transition risks and physic risks and this classification was used for the conceptualisation and operationalisation of the analysis categories in this research.

The paper examines sustainability reports (EU non-financial reporting) disclosures under the lens of TCFD Recommendations and tends to answer the following research questions:

RQ1: Do banks disclose transition risks and how?

RQ2: Do banks disclose physical risks and how?

RQ3: Do banks conduct scenario analysis and how?

The rest of the paper is as follows. Section 2 reviews the relevant literature and highlights the importance of disclosures for central banks, markets, and investors, the need for the use of new concepts of uncertainty (i.e., deep uncertainty) to cope with climate change, the sustainable data gap, the use of scenario(s)[9] as determinist and probabilistic approaches are insufficient, the relevance of text analytics for central banks analysis in a context of unstructured data growing. Section 3 analyses the methodology that has been used to conduct empirical research and it processes textual data with multivariate techniques through the use of software such as Iramuteq[10] and SAS Viya[11] combining unsupervised and supervised techniques. Section 4 describes the results bringing to public attention that despite a substantial step forward by the analysed banks in providing forward-looking and firm-specific measures (only 4 out of 12 banks give information on the scenario, transition, and physical risks in a comprehensive manner) one can conclude on the inadequacy, lack of transparency, lack of

[1] The United Nations Framework Convention on Climate Change (UNFCCC) was adopted at the 1992 Earth Summit in Rio de Janeiro and serves as a means for governments to deal with climate change. The ultimate objective of the convention is to stabilize greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous interference with Earth's climate (https://unfccc.int/).

[2] In 2020, the TCFD annual report dedicated an entire chapter to the topic of disclosures.

[3] Non-linearity is a relationship between climate variables and associated impacts that does not follow a constant or proportional progression; small variations in variables can cause disproportionately large or unpredictable changes. Some authors have dealt with non-linearity and the public perception of the implications of climate change (Hansen et al., 2012); others have dealt with "tipping points" that are examples of non-linearity and represent rapid and potentially irreversible changes in climate, the non-linearity and associated risks associated with the temperature targets set in the Paris Agreement have also been analysed.

[4] It was held in Paris, on 12 December 2015 and it entered into force on 4 November 2016.

[5] "Hovewer worlds leaders have stressed the need to limit global warming to 1.5°C by the end of this century […] because UN's Intergovernmental Panel on Climate Change indicates that crossing the 1.5°C threshold risks unleashing far more severe climate change impacts, including more frequent and severe droughts, heatwaves and rainfall" (https://www.consilium.europa.eu /it/policies/climate-change/paris-agreement/).

[6] Macro (or macroeconomic transmission channel are the mechanisms by which climate risk drivers affect macroeconomic factor, such as labour productivity and economic growth, and how these, in turn, may have an impact on banks through an effect on the economy in which banks operate. Macroeconomic transmission channels also capture the effects on macroeconomic market variables such as risk-free interest rates, inflation, commodities and foreign exchange rates (BCBS, 2021).

[7] Micro (or microeconomic transmission channels is a mechanism through which climate risk drivers affect banks' individual counterparties, potentially resulting in climate-related financial risk to banks and to the financial system. This includes the direct effect on banks themselves, arising from impacts on their operations and their ability to fund themselves. Microeconomic transmission channel also capture the indirect effects on name-specific financial asset held by banks (e.g., bond, single name CDS and equities (BCBS, 2021).

[8] The causal chains that explain how climate risk drivers give rise to financial risks that impact banks directly or indirectly through their counterparties, the assets they hold and the economy in which they operate (BCBS, 2021).

[9] In scenario analysis, mainly Integrated Assessment Models (IAMs), quantitative techniques and much less qualitative techniques are used, while the need to deal with qualitative and quantitative data together emerges.

[10] www.iramuteq.org

[11] www.sas.com

comparability, (also with respect to the pillars and recommendations of the TCFD) of the sustainability reports in the year 2021. Section 5 concludes the paper.

2. LITERATURE REVIEW

Climate change has inevitably to do with fossil fuel phase-down, climate finance, and more recently the goal of shifting to greater renewable energy capacity. Climate change will impact growth not only in the long term but also in the medium run (Nyambuu & Semmler, 2023). Few studies suggested that it may affect financial actors and the NGFS (2020) recommended the integration of climate-related risks into micro-prudential supervision setting out five recommendations and also the Bank of International Settlement (BIS) suggested to integrate climate-related and environmental risks into financial stability monitoring and micro-prudential supervision (BCBS, 2021).

Climate change, which conceptually differs from climate variability and global warming[12], is human-induced and according to Oxford Climate Society, the Intergovernmental Panel on Climate Change (IPCC) sequence of key findings about this statement follow these scientific results:

- broad overview of climate change science, discussion of uncertainties and evidence for warming (Houghton et al., 1990);
- the balance of evidence suggests a discernible human influence on global climate" (IPCC, 1995);
- most of the warming of the past 50 years is likely (> 66%) to be attributable to human activities (IPCC, 2001);
- warming is unequivocal, and most of warming of the past 50 years is very likely (90%) due to increase in greenhouse gases (Parry et al., 2007);
- it is extremely likely that human activities caused more than half the observed increase in global average surface temperature from 1951–2010 (Zhai, 2013).

Therefore, climate change (UNFCC) was defined as a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere.

This literature review is aimed to address the following research gaps: 1) sustainability report disclosures and central banks mandates, 2) sustainability data gap and solutions, 3) analysis of deep uncertainty for monetary policy and financial stability, 4) scenario analysis and related tools, and 5) text mining for central banks.

2.1. Disclosures and central banks' mandates

The first research gap in the literature on climate change and central banking revolves around the lack of analysis of the role of disclosures[13] and their relevance for central banks (BCBS, 2021; Siderius, 2023); they can be seen as a means for central banks to combine climate and environmental risks into financial stability (European Central Bank [ECB], 2021; Battiston et al., 2021; Ebner, 2018) monitoring and microprudential supervision (BCBS, 2020) and tackling climate as systemic risk (U.S. Governing Publishing Office Washington, 2021).

Conceptually, the topic of the disclosure can be organized into asymmetric, agency theory, or signal theory in the area of finance and corporate finance, while in the area of accounting and business administration, they tap into institutional theory and legitimacy theory.

In this literature review, we will focus on highlighting the need for greater homogeneity between different standards and frameworks.

The harmonization of sustainability reporting standards through the Corporate Sustainability Reporting Directive (CSRD) effective from 2024, the Sustainable Finance Disclosure Regulation (SFDR, 2019/2088), enforced in law from 10 March 2021, the European Taxonomy Regulation (2020/852) from 12 July 2020, and in particular the Implementing Technical Standards defined by the European Banking Authority (EBA) in January 2022, ITS (P3 ESG ITS), the requirements on climate disclosures with the International Sustainability Standard Board (ISSB) in the future, "should reduce the sustainability data gap within the European Union" (Banca d'Italia, 2021) and improve quality-quantitative aspects.

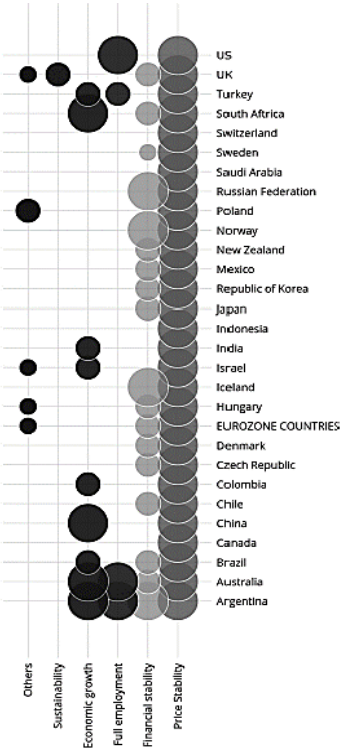
From the point of view of disclosure, central banks are required to asses financial institutions' performance and report how they account for environmental and social issues and need to identify the nature, persistence and magnitude of the climate-induced shocks central banks are expected to provide guidance and produce requirements (D'Orazio & Dirk, 2022; D'Orazio et al., 2023) on how banks are affected by climate change and how they impact the ecosystem with their activities; e.g., channeling financial flows to investment with low greenhouse gas emissions according to Paris Agreement[14].

This is related to the evolution of central banks' mandates (see Figure 1), which are heterogeneous in terms of objectives and mainly based on price stability. There is a need to reshape the traditional mandate of central banks alongside the two main objectives: monetary and financial stability.

The first central bank that innovatively updated its mandate was the Bank of England in 2021 by including in its objectives: sustainable and balanced growth that is also environmentally sustainable and consistent with the transition to a net-zero economy (The Bank of England, 2021).

[12] https://www.ipcc.ch/report/sr15/glossary/
[13] In general, as introduction to the whole topic (GARP, 2022).
[14] The opposite opinion pertains to "market neutrality" and central banks role (D'Orazio et al., 2023).

Figure 1. Structure of the Organization for Economic Co-operation and Development (OECD) and G20 central bank mandates



Source: D'Orazio et al. (2023).

2.2. Sustainability data gap, central banks, and climate change

The second research gap in the literature is the identification of sustainability data gaps in climate change and the use of an approach that gathers quantitative and qualitative data could improve statistical coverage, data quality, and scenario analyses.

There is substantial evidence that there are persisting problems of data availability, accessibility, usability, and reliability at the current time; this paper offers solutions for processing data with mixed variables (quali-quantitative) to combat the sustainability data gap and make a more affordable scenario.

Notwithstanding growing attention to the integration of risks connected to climate factors (Banca d'Italia, 2021) there is limited monitoring of climate risks (Banca d'Italia, 2021; Emambakhsh et al., 2022; Angelico et al., 2022)[15].

Emerging both scientifically and professionally is the need for central banks to gather qualitative and quantitative data through machine learning approaches and tools; the use of advanced techniques, properly documented, such as machine learning, could reduce the sustainable data gap, but still requires detailed information bases on which to calibrate the often unavailable algorithms, this article furnishes a solution and reflects on the use and limitations of classical inferential statistics.

[15] The same opinion is in the climate risk stress tests conducted by the ECB (2022) and are in line with an earlier survey of OMFIF (2020).

2.3. Deep uncertainty, monetary policy, and financial stability

The concept of uncertainty has taken several meanings in several fields but it can generally be defined as limited knowledge about the future (Walker et al., 2013)

Traditional uncertainty can be addressed with probabilistic models while deep uncertainty is characterized by limited available information, insufficient historical data, and variables interconnected in complex and not yet understood ways.

Climate-related financial risks are a concern for monetary policy and financial stability (Coueré, 2018; Schnabel, 2020) due to transition and physical risks and their direct and indirect effects (Aversa et al., 2022), they cause asset stranding (Ansar et al., 2013) that could lead to divestment by rational investors, to an adverse effect on employment, tax revenues, and trade dynamics and to a possible "climate minsky moments".

Central banks are revisiting their backward-looking risk models in favor of the integration of forward-looking climate-related risks, at the same time the current climate-economy models are limited in a context of radical uncertainty, therefore climate scenario analysis should be performed on systems-based approaches applying non-equilibrium models and more qualitative tools, e.g., those that blended socio-technical variable (Svartzman et al., 2021; Aversa, 2023a).

Deep uncertainty, which refers to when deterministic and probabilistic approaches are insufficient for representing future states, and the consideration of multiple plausible futures (scenarios) is required (McPhail et al., 2020) as a new concept of investigation.

Likewise, Lars Peter Hansen has written that the longer-term notions of uncertainty, including potential climate change, making confident probability statements is at best challenging and arguably contrived and these probabilities often leads to some well-warranted skepticism consequently, a researcher must push beyond the traditional notion of risk and uncertainty like the knigthian one[16].

The need to use new tools for analysis and monetary policy emerges, and also the need to use other tools stems, according to Hansen, from at least three factors:

1) historical data for quantifying climate-change uncertainty are limited and typical risk-based analyses presume rich historical evidence to infer expect return[17] reliably;

2) research to date has done little to isolate the impact for different sources of climate-change uncertainty;

3) using model based on rational expectations to estimate probabilities face the problem that expectations are necessarily subjective in the case of climate economics.

To aid central banks to handle deep uncertainty in climate change economics must be implemented the use of scenario.

[16] Risk refers to a "particular type of uncertainty that includes a confident knowledge of probabilities".

[17] Rational expectations hypothesis has to "figure out credible probabilities for a good approximation".

2.4. Scenario analysis and central banks

The scenario is "a plausible description of how the future may develop based on a coherent and internally consistent set of assumptions about key driving forces (e.g. rate of technological change, prices) and relationships. Scenarios are neither predictions nor forecasts, but are used to provide a view of the implications of development and actions" (BCBS, 2021, p. 5). Scenario analysis is "a tool that is used to enhance critical strategic thinking. A key feature of the scenario analyzed is to explore alternatives that may significantly alter the basis for "business-as-usual" assumptions. Accordingly, they need to challenge conventional wisdom about the future" (BCBS, 2021, p. 6).

According to Tol (2023), scenario analysis is a conditional prediction and if the description of future events is incomplete, as it is in most cases and certainly in climate change, predictions are necessarily conditional. Alongside quantitative scenario analyses, qualitative tools and techniques should be used to combine qualitative and quantitative variables through machine learning.

2.5. Central banks and text analytics

Text analytics (Giuliano, 2004; Tipaldo, 2014; Kalamara et al., 2020) also named text mining, natural language processing or computational linguistics is valuable in addressing relevant topics for central banks (The Bank of England sets out this new approach to data in its Bank's Strategic Plan).

As a tool less used in economic analysis, it might be seen as a means of non-orthodox economics, a source of data with a potential that "has not been fully tapped" (The Bank of England, 2015), serves to communicate the measurement of uncertainty in terms of future policy directions (Carney, 2013) with a consistent message.

The Bank of England (2015) and Bholat (2015) argued text mining contributes to understanding the interaction between monetary, macro-prudential and micro-prudential policy and it adds new weapons to central banks' analytic arsenal (Hansen et al., 2018).

3. RESEARCH METHODOLOGY

The methodology of the research lays on unsupervised and supervised learning with a complex, multi-stage, and iterative process that performs the investigation on the matter more efficiently and effectively. The research techniques applied data mining for textual data also named text analytics and used unsupervised learning to extract meanings, patterns, and structure hidden in unstructured textual data. It also applied specificities and lexical correspondence analysis, similarities, and cluster analysis with Reinert's (1990) method.

Reinert's method also known as Alceste, a method of hierarchical decedent analysis, creates grouping classes of statements that are as homogeneous within themselves and as different from each other as possible. In this way, Alceste determines the specific vocabulary of each class, consisting of the words that are significantly more present in the statements of the class than in the rest of the corpus (Reinert, 1990).

The purpose of the research relates to the analysis of the disclosure information within the sustainability reports (Directive 2014/95/EU) of the listed banks on the FTSE Italia All-Share index of Borsa Italiana.

The research, made with quantitative tools for qualitative analysis (data mining for textual data), approached the topic with mixed research methods and design.

The corpus is made of 12 banks of which seven are included in the FTSE MIB index, four in the FTSE Italia Mid Cap, and 1 in the FTSE Italia Small Cap, these indices constitute the three subsets of the FTSE Italia All-Share macro-aggregate.

Table 1. FTSE Italia All-Share banks

Bank name	*Index*	*Coding*	*Year*
UniCredit Bank	FTSE MIB	****001 *bank_unicredit *index_ftsemib *year_2021	2021
Mediobanca	FTSE MIB	****002 *bank_mediobanca *index_ftsemib *year_2021	2021
Intesa San Paolo Bank	FTSE MIB	****003 *bank_intesasanpaolo *index_ftsemib *year_2021	2021
Banca BPER	FTSE MIB	****004 *bank_bper *index_ftsemib *year_2021	2021
Banca BPM	FTSE MIB	****005 *bank_bpm *index_ftsemib *year_2021	2021
Banca Mediolanum	FTSE MIB	****006 *bank_mediolanum *index_ftsemib *year_2021	2021
Banca Generali	FTSE MIB	****007 *bank_generali *index_ftsemib *year_2021	2021
Banca Monte dei Paschi di Siena	FTSE Italia Mid Cap	****008 *bank_mps *index_ftseitaliamidcap *year_2021	2021
Banca Popolare di Sondrio	FTSE Italia Mid Cap	****009 *bank_popolaresondrio *index_ftseitaliamidcap *year_2021	2021
Banca Credito Emiliano	FTSE Italia Mid Cap	****010 *bank_creditoemiliano *index_ftseitaliamidcap *year_2021	2021
Banca Credito Valtellinese	FTSE Italia Mid Cap	****011 *bank_creditovaltellinese *index_ftseitaliamidcap *year_2021	2021
Banca di Desio e Brianza	FTSE Italia Small Cap	****012 *bank_desiobrianza *index_ftseitaliasmallcap *year_2021	2021

The sustainability reports of 12 banks (partition variables) belong to the FTSE Italia All-Share index for the year 2021.

The software IRaMuTeQ 0.7 alpha 2 has been used for the text multidimensional analysis. IRaMuTeQ (*Interface de R pour les Analyzes Multidimensionnelles de Textes et de Questionnaires*) developed by the French programmer Pierre Ratinaud is an open source, multifunctional and statistic software for textual data analysis based on Python and also on statistical software R (Souza et al., 2018).

The contents were saved in text format (UTF-8 coding) for automatic analysis, each text was accompanied by the text partition key variables indicated below:

a) identifier: **** number
b) bank: * bank_bank name
c) index: * index_index name
d) year: * year_number

Example of a trace: **** 001 * bank_unicredit * index_ftsemib * year_2021.

As written, the techniques used are text analytics and combine unsupervised learning with techniques of information extraction and information retrieval.

The conduct of the research follows a multi-stage iterative process using four methods of analysis:

- Descriptive analysis: type-token ratio (TTR), TTR made using lemmas (L-TTR), Giraud Index and Herdan Index, Zipf's curve and its slope to verify lexical richness and justify the automatic and semi-automatic treatment of the data;
- Specificities analysis and correspondence analysis;
- Similarities analysis;
- Cluster anaysis with Reinert's method;

3.1. Data and lexicometric measures

The treatment of the qualitative data of the collective of documents is carried out with multivariate techniques, it begins with the lexicographic analysis (lexicometric measures) of the corpus (homogeneous, monolingual Italian and referred to the time period 2020–2021), subsequently the analysis of specificity and lexical correspondences, similarity and cluster analysis with the Reinert's method.

The phases of construction and validation of the corpus are the following two:

a) lexicalization (to identify complex and compound lexias);

b) application of multivariate techniques after lexicalization.

In phase (a), the lexicographic analysis carried out with Iramuteq, before the lemmatization and after the lexicalization, identifies 12 texts in the corpus, 99,617 tokens, 9,065 types and a number of hapaxes equal to 3,848 in absolute terms, corresponding to 3.86% of occurrences and 42.45% of forms (Table 2).

The descriptive statistics are summarized in the following abstract.

Table 2. Corpus abstract after the lexicalization

Parameters	*Values*
Number of texts	12
Number of text segments	2,476
Number of occurrences	99,617
Number of forms	9,065
Number of hapaxes	3,848 (42.45% of forms – 3.86% of occurrences)

Source: Author's elaboration using Iramuteq.

In phase (b) the lexicalization generated 25 lexias between compound and complex, the lexias are sets of two or more words that take on a different meaning (Table 3), and the terms taken into account have a maximum frequency of 72 and minimum frequency 15.

Table 3. Lexicalization

Lexicalized word		*Lexicalized word*		*Lexicalized word*		*Lexicalized word*		*Lexicalized word*	
1. bilancio integrato	72	1. sviluppo sostenibile	36	1. rischi climatici	26	1. climate change	18	1. impatti ambientali diretti	16
2. impatti ambientali	54	2. impatto ambientale	33	2. fattori esg	25	2. sistema di gestione	18	2. rischi climatici e ambientali	16
3. gestione dei rischi	46	3. dichiarazion consolidata	30	3. dichiarazion consolidata	23	3. diritti umani	17	3. finanza sostenibile	15
4. gruppo banco desio	39	4. carattere non finanziario	27	4. codice etico	22	4. strumenti finanziari	17	4. riduzione degli impatti	15
5. prodotti e servizi	37	5. linee guida	27	5. consiglio di amministrazione	19	5. governo societario	16	5. risk management	15

The lexicalization highlights, with the words "*integrated financial statements*", "*consolidated statements*" and "*non-financial character*", the aspects of compliance with Directive 2014/95/EU.

The focus is on environmental risk and its management, its direct and indirect impacts together with ESG criteria, and, the reference legislation that governs the rules of Integrated Reporting (IR).

In the analysis conducted on these documents, compared to those of the year 2020, a particular focus emerges on the risks connected to the climate: two lexias originate "*climate risks*" and "*climate and environmental risks*", respectively with absolute frequency values of 26 and 16.

Further descriptive measures of the corpus consist of the indices indicated below and depicted in Table 4:

- TTR;
- Index Guiraud;
- Index Herdan;
- TTR after lemmatization;
- Zipf's curve;
- Zipf's slope curve;
- Abstract after lemmatization and L-TTR.

Table 4. Lexicometric measures

Lexicometric measures	*Formula*	*Value*
TTR	$(V/N) * 100$	$(9065/99617) * 100 = 9.09\%$
% Hapax	$(V_1/V) * 100$	$(3848/9065) * 100 = 42.45\%$
Zipf's slope curve	$(\log N/\log V)$	$(\log 99617/\log 9065) = 11.5/9.11 = \lvert 1.26 \rvert$
Guiraud index	$V/\sqrt{N}$	$9065/\sqrt{99617} = 9065/315.62 = 28.72$
Herdan index	$(\log V/\log N)$	$\log 9065/\log 99617 = 3.95/4.99 = 0.79$

Type-token ratio (TTR): The ratio between the number of different words (type) and the total number of words (occurrences or token) is characterized by a value considerably lower than 20%, amounting to 9.09% indicates a solid validity of the lexicometric measure and allows to consider the corpus susceptible to an automatic or semi-automatic treatment.

The TTR is therefore a ratio between the width of the vocabulary (*V*) of 9,065 units and the size of the corpus (*N*) of 99,617 occurrences[21]: it represents the lexical extension. Later in the analysis, the version based on lemmas *(L-TTR) = Lemma/N* was also used.

Hapax percentage: The absolute number of hapaxes is 3.848 corresponding to 3.86% of occurrences and 42.45% of forms, its determination which is less than the 50% limit threshold leads to the evaluation of the correct size and width of the corpus and the correct processability through text analytics.

Guiraud index: The corpus continues to be validated using the Guiraud index, the ratio between the number of forms ($V = 9{,}065$) and the square root of the occurrences ($\sqrt{N} = 315.62$), it indicates a considerable lexical richness since its value is 28.72 in absolute terms[22].

Herdan index: Herdan index, expressed by the ratio between the logarithm of the types ($\log V = 3.95$) and the logarithm of the occurrences ($\log N = 4.99$) confirms the considerations relating to lexical richness.

Abstract (after lemmatization) and L-TTR: The lemmatization is the analysis phase that brings the words back to the canonical form, that is how they are represented in the dictionary of the language, in this analysis it is characterized by the improvement of the Hapax percentage and the TTR which are reduced further:

- Number of texts: 12;
- Number of occurrences: 99,617;
- Number of forms (lemmas): 6.311;
- Number of hapaxes: 2,417 (2.43% of occurrences - 38.30% of forms);
- The mean of occurrences by text: 8,301.42.

$$L - TTR = \frac{V}{N} * 100 = \frac{6{,}311}{99{,}617} * 100 = 6.33\% \quad (1)$$

The L-TTR index which is equal to 6.33% clearly improves the TTR (reducing the value of the numerator generates a better and smaller result).

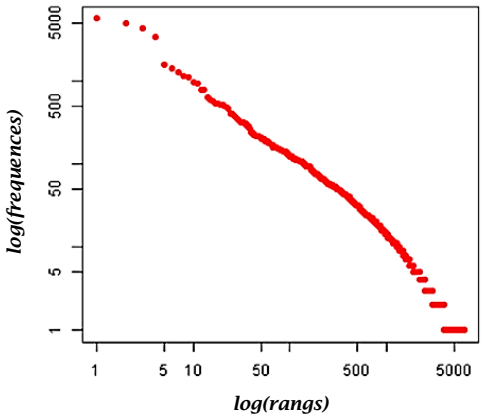
Zipf's curve and Zipf's slope curve: Words in a dictionary always have a well-defined distribution in terms of occurrences. Its form is known and was identified by the linguist George Kingsley Zipf[23] (Zipf's law) who observed the following relationship between the rank and frequency of a word:

$$F * R = c \quad (2)$$

where, *F* indicates the frequency and *R* the rank.

Frequency is inversely proportional to rank — the higher the rank the lower the frequency as seen in the graph.

Figure 2. Zipf's curve



Source: Author's elaboration using Iramuteq.

The Zipf's Curve, expressed by the logarithm of the rank on the abscissa by the logarithm of the frequency on the ordinate, takes the form in Figure 1: the higher the rank, the lower the frequency.

This is a rule to be understood in a statistical sense on average because it is necessary to take an average value of occurrence of the words belonging to a certain neighbourhood of the considered rank (there are not all the possible frequencies and then there are the *ex aequo*).

Therefore, it is more appropriate to express the formula as follows:

$$f * r^a = c \quad (3)$$

which on a logarithmic scale becomes:

$$a \log(r) + \log(f) = c \quad (4)$$

the report can be written:

$$\log(f) = c + a \log(r) \quad (5)$$

Zipf's slope curve: The coefficient of Zipf's Curve is the slope of a line on a graph with logarithmic coordinates: the abscissa (x-axis) indicates the logarithm of the rank and the ordinate (y-axis) the logarithm of the frequency representing the proportion of different words in a vocabulary (richness lexical).

$$Zipf's\ Curve\ Slope = \frac{logN}{logV} = \frac{log9{,}065}{log99{,}617} = \frac{3.95}{4.99} = |1.26| \quad (6)$$

The slope, on the logarithmic coordinate graph, is well approximated by *logN/logV* and its optimal value must be around 1.3 in absolute terms, in this case, it has a value of |1.26|.

The statistics examined summarize the possibility of subjecting the corpus to automatic or semi-automatic processing, the statistical-quantitative characteristics are robust for validation and lexicometric treatment of text data.

The study of the distribution of words in the text is carried out with a table and a Pareto chart ordered according to the decreasing frequencies of

[21] It is an index sensitive to the size of the corpus; its limitation lies in the fact that as the number of occurrences increases its value tends to decrease and therefore to fall below the 20% threshold due to the fact that the graphic forms, they tend to repeat themselves. As mentioned, if its value is less than 20% the corpus is considered adequate for a lexicometric type treatment.

[22] The minimum threshold of the Guiraud index is 22.

[23] Zipf's Law provides a probability distribution for the frequency of words in text. It is like a discrete version of the Pareto distribution. A feature of Zipf's law is that a plot of the frequency of words versus the rank of the word on a log scale will be approximately linear. Perl can be used to tabulate the frequencies of words in a document or database to see if they follow Zipf's law.

the terms together with the indication of the frequency bands of the words (high, medium, and low).

The high-frequency range starts from the frequency value 1.263 up to 315, the average one is between 314 and 88 including the keywords, while the low one has a range of values from 64 to 2.

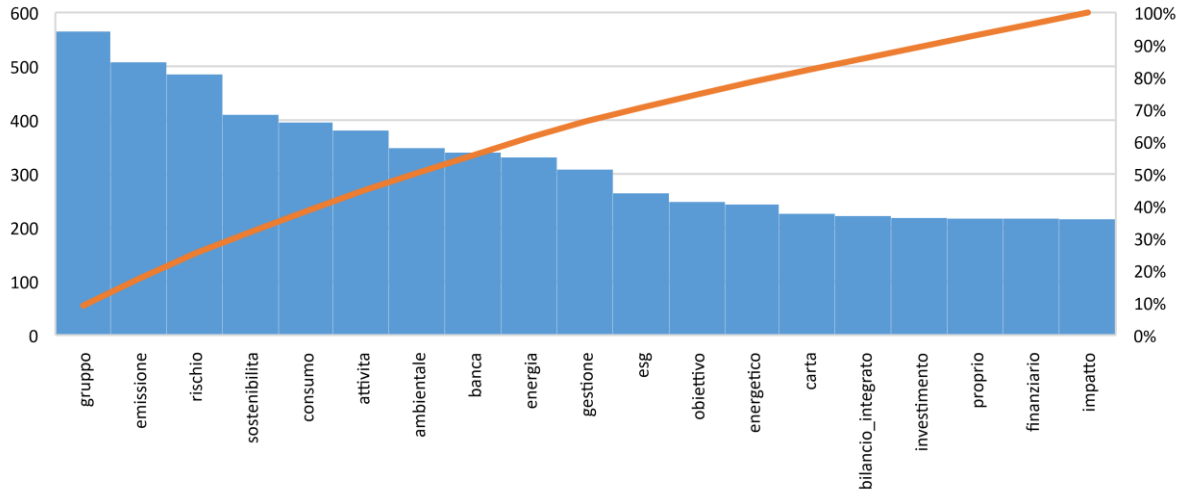
The high-frequency range includes the topic words: the names "*group*", "*issue*", "*risk*", "*sustainability*" and "*consumption*" constitute the first 5 terms to be considered, overall, it shows the effort of banks toward an economy with low environmental impact.

Table 5 depicts the frequency of the first 19 words starting from the highest frequency to the lowest and using the Pareto chart (Figure 3) for graphical synthesis.

Table 5. Frequency of the first nineteen words

Words (Italian/English)		*Count*	*Part of speech*
gruppo	group	565	nom
emissione	emission	508	nom
rischio	risk	485	nom
sostenibilita	sustainability	410	nom
consumo	consumption	396	nom
attivita	activity	381	nom
ambientale	environmental	348	adj
banca	bank	340	nom
energia	energy	331	nom
gestione	management	308	nom
esg	esg	264	nr
obiettivo	objective	248	nom
energetico	energetic	243	adj
carta	paper	226	nom
bilancio_integrato	integrated_budget	222	nr
investimento	investment	218	nom
proprio	own	217	adj
finanziario	financial	217	adj
impatto	impact	216	nom

Figure 3. Pareto graph (frequency of the first nineteen words)



The term document matrix includes reference words aimed at the effort towards a low-carbon economy by reducing direct and indirect impacts on the climate changes determined by company activity.

The companies analysed in the year 2021 communicate, similarly to the sustainability reports of 2020, the strategy adopted for the transition processes towards a low-emission economy through the reduction of greenhouse gases issued.

The words with the greatest representation indicate that the first 5 items — excluding the term "*group*" — are "*emission*", "*risk*", "*sustainability*", "*consumption*", and "*activity*", they support the transition to a net zero emissions economy over time (The Bank of England, 2022).

The analysis of the co-occurrences helps to confirm this assessment.

****001 *bank_unicredit *index_ftsemib *year_2021

esposizione al settore delle energie rinnovabilicomplessivi generati sul mercato obbligazionario esg di asset under management convertiti in investimenti esg zero **emissioni** delle nostre attività entro il 2030 **emissioni** del nostro portafoglio entro il 2050

In fact, emissions are related to renewable energy, products for customers, and investments in the portfolio for achieving the net-zero strategy and for alignment with international standards.

****010 *bank_credem *index_ftseitaliamidcap *year_2021

scenario ottimistico di contrasto efficace al cambiamento_climatico e riduzione significativa delle **emissioni** di gas ad effetto serra nell atmosfera scenario pessimistico comunemente associato all espressione business as usual o nessuna mitigazione in_cui la crescita delle **emissioni** continua ai ritmi attuali

The emissions also concern the forecasts of future scenarios (scenario analysis) such as the business as one usual in which the growth of issues is continuous at the current pace according to the expression of Banca Credito Emiliano.

****008 *bank_mps *index_ftseitaliamidcap *year_2021

come già espresso anche nelle dichiarazione non finanziaria dell anno 2020 il gruppo intende inoltre focalizzare l attenzione sul miglioramento degli impatti diretti in termini di **emissione** di co2 derivanti dal consumo di gas naturale attualmente rappresentanti mediamente il 75 delle **emissioni** dirette totali del gruppo

It is clear that direct and indirect emissions, with particular reference to carbon dioxide, are subject to disclosure in sustainability reports.

****001 *bank_unicredit *index_ftsemib *year_2021

passaggio a un economia a basse **emissioni** di carbonio promozione di strumenti_finanziari sostenibili finanziamento dell efficienza energetica

Likewise, a "*low-emission economy*" must be accompanied by sustainable financial instruments and financing for energy efficiency.

****010 *bank_credem *index_ftseitaliamidcap *year_2021

nel 2021 il gruppo credem ha quantificato per la prima volta una stima delle **emissioni** di co2 indirette scope_3 collegate a finanziamenti e investimenti

In the quantification of emissions, according to the GHG Protocol, information is given for those of Scope 1, Scope 2, and for those of Scope 3 not all banks are able to supply; this aspect is in disagreement with the pillars and TCFD recommendations while Banca Credito Emiliano is perfectly compliant with respect to the metrics required by this framework.

****012 *bank_desiobrianza *index_ftseitaliasmallcap *year_2021

le fonti delle **emissioni** classificate_come scope1 sono generalmente possedute e controllate direttamente dal gruppo scope_2 il relativo indicatore considera le **emissioni** derivanti dalla produzione di energia elettrica acquistata dal gruppo e consumata per le apparecchiature elettriche e l illuminazione all interno degli edifici

****005 *bank_bpm *index_ftsemib *year_2021

il costo inoltre è interamente ribaltato sull utilizzatore tale scelta risulta coerente con le indicazioni del ghg protocol che secondo l operational control approach vedrebbero le **emissioni** di scope_2 derivanti dai consumi di_questo immobile escludibili dal perimetro delle **emissioni** dell organizzazione

It is useful to specify that Scope 1 emissions are those owned and controlled by the group, those of Scope 2 are indirect and those of Scope 3 derive from collaboration with third parties.

****011 *bank_creditovaltellinese *index_ftseitaliamidcap *year_2021

implementazione dei processi digitalizzati che ha favorito un aumento della dematerializzazione e del risparmio di carta contribuendo alla riduzione delle **emissioni** di tonnellate di co2 equivalenti

Among the measures aimed at reducing emissions, those connected with dematerialization and the use of certified paper that feeds circular economy processes are worthy of note.

The analysis conducted in 2021 with reference to the word risk offers interesting arguments with respect to the research hypotheses.

Regarding transition risks, Banca Credito Emiliano provides information on physical risks and integrates climatic aspects into corporate strategy and governance, giving centrality to this type of risk alongside traditional ones: credit, market, operational, and reputational. In addition, physical and transition risks have been included in the valuation of the loan, investment, and asset management portfolios.

Banca UniCredit denotes awareness of how policy and legislation, investors, regulatory bodies, and stakeholders require transparent information regarding the impacts caused by climate change and climate risk (the company has organized a team of sustainable finance advisers).

****010 *bank_credem *index_ftseitaliamidcap *year_2021

nonché sulla rilevanza dei **rischi** fisici e di transizione sui portafogli crediti investimenti e immobili definizione di_una roadmap di intervento con l obiettivo di consentire al gruppo di allinearsi alle suddette aspettative integrando gli aspetti climatici nella strategia e governance aziendale

****010 *bank_credem *index_ftseitaliamidcap *year_2021

nonché nel framework e nei processi di gestione_dei_rischi implementazione della roadmap e declinazione operativa degli interventi 2 3 **rischio** climatico il cambiamento_climatico sta guidando alcune delle più profonde trasformazioni globali gli impatti diretti e indiretti legati al cambiamento_climatico sono già visibili

****010 *bank_credem *index_ftseitaliamidcap *year_2021

nonché in termini di cambiamenti a livello di policy e legislazione investitori enti regolatori e altri stakeholder richiedono sempre più spesso informazioni trasparenti relativamente agli impatti determinati dal cambiamento_climatico sulle imprese e viceversa il **rischio** climatico include

****001 *bank_unicredit *index_ftsemib *year_2021

il ruolo del sustainable finance advisory team maggiori informazioni sono disponibili nel capitolo capitale naturale del bilancio_integrato maggiori informazioni sono disponibili nella sezione gestione del **rischio** e compliance del supplemento e nel capitolo gestione del **rischio** e compliance aumentare il coinvolgimento

****010 *bank_credem *index_ftseitaliamidcap *year_2021

i **rischi** posti dal cambiamento_climatico potrebbero materializzarsi attraverso effetti negativi sulle categorie di **rischio** tradizionali già oggetto di presidio specialistico da parte del gruppo quali il **rischio** di credito di mercato operativo e reputazionale

****001 *bank_unicredit *index_ftsemib *year_2021

il quadro di riferimento del settore finanziario per identificare valutare e gestire il **rischio** ambientale e sociale dei progetti sustainable finance advisory team unicredit si è data nuovi obiettivi ambiziosi per ridurre l impatto_ambientale delle proprie attività sia operative_che creditizie

****010 *bank_credem *index_ftseitaliamidcap *year_2021

per esprimere una prima valutazione sul potenziale **rischio** climate_change in termini di **rischi** fisici e di transizione sui portafogli inerenti i finanziamenti e gli investimenti di proprietà e gestioni patrimoniali

Mediobanca in relation to the assessment of risks relating to climate change proceeds with their identification according to the approach proposed by the TCFD.

****002 *bank_mediobanca *index_ftsemib *year_2021

valutati i **rischi** relativi al cambiamento_climatico tali **rischi** sono_stati identificati secondo l approccio proposto dalla task force on climate related financial disclosures tcfd facendo una distinzione tra quelli fisici e quelli di transizione che sono risultati più rilevanti per il gruppo

Banks enhance the new types of risks alongside traditional ones (credit, market, operational, and reputational) also in Banca Intesa San Paolo.

****003 *bank_intesasanpaolo *index_ftsemib *year_2021

rischio di credito **rischio** di mercato **rischio** operativo e si sviluppa in stretta integrazione con il presidio dei **rischi** reputazionali valorizzando le interconnessioni che caratterizzano questi profili di **rischio**

****012 *bank_desiobrianza *index_ftseitaliasmallcap *year_2021

un accensione continuativa anche nelle ore notturne per garantire il ricambio d aria e ridurre il **rischio** di propagazione del contagio del virus

The lower presence of personnel due to COVID-19 implies less use of heating and cooling systems, resulting in a reduction in consumption (but this does not happen for all banks as some underline how the nighttime switching on of ventilation systems — imposed by anti-COVID regulations — led to increased energy use and increased direct emissions.

****012 *bank_desiobrianza *index_ftseitaliasmallcap *year_2021

oltre a promuovere modelli di **consumo** delle risorse più consapevoli e_quindi la possibilità di poter anche contenere i costi operativi l europa in particolar modo si è prefissata di darsi degli obiettivi e delle scadenze_che per quanto ovvio debbano riportare risultati controllati e misurati in_modo certificato

****004 *bank_bper *index_ftsemib *year_2021

anche sul dato relativo ai kwh risparmiati pesa il calo dei **consumi** causato dal periodo pandemico il gruppo bper all interno del nuovo piano industriale ha redatto il nuovo piano energetico che prevede target di riduzione delle emissioni in linea con gli obiettivi dell accordo di parigi scenario 2

****008 *bank_mps *index_ftseitaliamidcap *year_2021

approvvigionamento di energia da fonti rinnovabili monitoraggio continuo dei dati di **consumo** diagnosi energetiche su immobili clusterizzati l utilizzo efficiente e socialmente responsabile dell energia è particolarmente significativo per il controllo degli impatti complessivi del gruppo sull ambiente

Consumption refers to the various sources of energy used: traditional and renewable but also water and the supply of raw materials such as paper (certified paper), they are included in the reports (smart working has decreased the use of paper).

The consumption of resources also refers to more aware circular economy practices such as more efficient management of consumption and separate waste collection.

For the research questions of our study, there are reports aimed at mitigating the impacts of climate change through the reduction of greenhouse gases in the atmosphere.

In the matrix of terms for documents analysed, a greater orientation appears to provide information consistent with the research hypotheses (we will see later that the information extraction techniques refer greater attention to transition risks, physical risks, and scenario analysis).

3.2. Specificities and lexical correspondence analysis

The analysis of the specificities allows us to verify if a word is specific or not of a single text/category of texts.

A term is considered "specific" if it appears in a document to a greater extent than expected (the reverse indicates a rare word).

It is obtained through the comparison between the observed frequency and the expected frequency of a word within a document, and a calculation of probabilities based on hypergeometric distribution specificities is identified.

The tables show the positive specificities (with a value greater than 2) of all the banks referring to the first 20 words found in the "specificities table" by Iramuteq (this way of representing data facilitates reading and understanding).

The specific words sorted by decreasing score identify the areas of greatest interest for each bank.

The interpretation takes place both by reading the single words and through an overall look and comparison with the terms and other variables investigated.

Table 6. Specificities (BPER, BPM, and Credito Emiliano)

X.bank_bper		*X.bank_bpm*		*X.bank_creditoemiliano*	
bper	102.3505	rischio	19.7605	credem	18.00
scenario	20.787	eligible	16.8631	riconducibile	9.10
tco2e	17.1065	green	16.8462	spreco	7.89
transizione	16.8339	bpm	16.0868	inerente	7.00
settore	12.7806	total	10.9905	fisico	6.10
portafoglio	11.5752	asset	10.325	cass	5.98
high	11.4998	mln	9.859	foresta	5.62
ftv	9.2689	transizione	9.568	emissione	5.26
intensity	9.2689	esposizione	6.9474	edificio	5.06
footprint	9.1528	creditizio	6.625	transferta	5.03
fatturato	8.5352	fisico	6.4889	collegare	4.79
macro	8.426	neutrality	5.5052	percorrenza	4.65
medio	8.2116	tassonomia	5.3677	livello	4.60
tcfd	7.697	quota	4.6689	working	4.56
fabbricazione	7.4993	ue	4.3577	non-sono	4.44
rating	7.4203	report	4.23	serra	3.97
alto	6.8639	alto	3.9284	investimento	3.90
orizzonte	6.6284	politico	3.8955	ove	3.85
stima	6.5882	opportunita	3.8715	albero	3.71

- *Banca BPER* uses the word "*scenario*" as a specific word, and connotes itself as a financial institution of interest to verify the hypothesis of studies, in particular together with UniCredit and Mediobanca.

In the text there are also "*physical risk*" and "*transition*" and their association with portfolios, credits, investments, and financing.

The specific word "tco2" defines the unit of measurement adopted to report the reduction of carbon dioxide.

Banca BPER identifies the word risk very well and makes a portfolio assessment of it, speaking of scenario analysis and calling into question the TCFD. The reference scenario is the business as one usual (BAU) "in which a strong increase in temperature and climatic, physical and transition risk scenarios are assumed", and forecasts are determined that alter the base scenario.

Banca BPER makes a specific reference to the TCFD benchmark when compiling the sustainability report.

- *Banca BPM* encountered the terms "*risk*" and "*transition*" as positive specific features "physical" and "*taxonomy*" (European), the latter derives from Regulation 2020/852 EU and introduced a common classification framework for eco-sustainable activities or a guide for investing in economic activities that contribute towards the transition to a lower carbon economy.

******005 *bank_bpm *index_ftsemib *year_2021**

per tale ragione abbiamo analizzato tutti gli eventi di rischio acuto e cronico associati al cambiamento_climatico e mappati dal regolamento sulla **tassonomia** ue individuando quelli più rilevanti per le regioni italiane in_cui operiamo ovvero i rischi di inondazione fluviale etc

******005 *bank_bpm *index_ftsemib *year_2021**

esg linked loan o sustainable linked loan progetti coerenti con la **tassonomia** ue quali ad_esempio operazioni che beneficiano della garanzia sace green new deal finalizzati alla produzione di energia da fonti rinnovabili nel campo della mobilità sostenibile e dell economia circolare

This information constitutes an aspect linked to mandatory rules and the indication of adequate information to stakeholders.

- *Banca Credito Emiliano* identifies waste and emissions among its peculiarities to rationalize the reduction of pollution activities with reference to the mobility of the workforce, to the efficiency of owned facilities.

The next set analyzed consists of the group in Table 7.

Table 7. Specificities (Credito Valtellinese, Desio e Brianza, and Generali)

X.bank_creditovaltellinese		*X.bank_desiobrianza*		*X.bank_generali*	
elettrico	10.8045	appendix	39.7106	generale	48.4294
energia	9.826	gruppo_banco_desio	38.9874	banca	13.1897
impianto	7.9907	integrita	37.1859	sede	12.0188
consumo	7.984	highlights	34.4768	acqua	10.4547
fonte	7.0402	opinion	34.4768	sga	7.0986
teleriscaldamento	6.292	lettera	29.9994	ricorso	6.4757
mix	6.1109	responsabilita	28.4905	consumare	5.4504
milano	5.835	sostenibilita	27.2038	consulente	5.2108
rinnovabile	5.3866	desio	23.0889	milano	4.3258
peso	5.1292	etico	22.0879	teleriscaldameento	3.7679
gasolio	4.9843	profilo	19.2865	rifiuto	3.6906
totale	4.9532	coefficiente	19.0751	ristrutturazione	3.6661
benzina	4.8678	indico	15.5844	operativo	3.3915
fsc	4.7614	metodologia	14.6801	pubblico	3.2023
condominiale	4.7165	gri	14.0919	mobilita	3.1351
produzione	4.1003	bilancio	12.8814	conversione	3.1062
riscaldamento	4.0473	effettivo	11.5908	mezza|mezzo	2.8063
associare	4.0042	governance	11.2489	rispetto_all	2.5424
sito	3.8579	precedente	9.638	naturale	2.5332

- *Banca Credito Valtellinese* shows words like "*consumption*", "*plant consumption activity*", "*methods of using energy*", "*the use of renewables*", and "*heating emissions*" are used as specific characteristics.
- *Banca Desio e Brianza* has the specificities connected to the reporting methodologies of the activities aimed at reducing pollution and to an ethical-value reference to the environmental issue.
- *Banca Generali* displays the presence of the terms "*consume*" and "*waste*" which refer to good environmental practices.

Table 8 addresses the specificities of Banca Intesa San Paolo, Mediobanca, and Mediolanum.

Table 8. Specificities (Intesa San Paolo, Mediobanca, and Mediolanum)

X.bank_intesasanpaolo		*X.bank_mediobanca*		*X.bank_mediolanum*	
esg	49.351	mediobanca	16.8	mediolanum	132.9384
divisione	23.3717	consumo	15.5827	fondazione	46.1152
sanpaolo	22.0474	scopo	13.4254	bambino	21.4367
former	16.9926	rifiuto	10.7861	progetto	13.5052
reputazione	12.861	carta	10.4868	family	13.0947
persona	11.477	sede	8.3014	kit	13.0947
rischio	10.5273	unita	8.2187	ong	11.9036
intendere	9.7656	utilizzo	6.6749	onlus	11.7078
gestione_dei_rischi	9.5614	superficie	6.1337	banker	11.2555
codice_etico	9.2835	rinnovabile	5.8185	evento	10.9683
formazione	8.822	utilizzare	5.387	spagna	10.8884
operativita	8.8024	smaltimento	5.103	donazione	9.0241
comitato	8.7595	stimare	4.8901	collettivita	7.1733
inclusione	8.0392	idrico	4.8408	favore	7.1733
governo	7.6174	condominiale	4.8316	culturale	7.1686
collaboratore	7.5246	disponibile	4.6207	associazione	7.0698
tematica	7.1044	pericoloso	4.5472	banco	5.8752
relazione	6.9562	sostituzione	4.5472	stampa	5.255
conservazione	6.9094	metodo	4.517	raccolta	5.1531

- *Banca Intesa* is particularly connected to the issues of risk, risk management, ethical implications, and inclusion for the transition from a carbon-intensive economy to one with a low environmental impact.
- *Mediobanca* uses the terms circular economy practices with the words "*waste*", "*paper*", "*disposal*", and "*water*" and choices related to renewable energy.
- *Banca Mediolanum* mainly exhibits the practices with positive specificities: "*foundation*", "*child*", and "*project*" giving ample information on its social practices.

Table 9. Specificities (Monte dei Paschi di Siena, Popolare di Sondrio, and UniCredit)

X.bank_mps		*X.bank_popolaresondrio*		*X.bank_Unicredit*	
dichiarazione	16.3616	environment	25.0124	bilancio_integrato	238.8604
mps	14.992	carattere_non_finanziario	24.762	inicredit	98.3464
dimostrane_che	13.9204	dichiarazione_consolidata	20.4883	modalita	51.5564
banca	10.1166	investimento	11.3408	gestione	31.0952
consolidato	9.5706	provento	10.4749	topic	30.9044
attuazione	9.4537	bond	9.6254	gri	29.3228
obiettivo	8.0802	green	9.2336	spiegazione	27.7033
gasolio	8.0225	proprio	7.4789	perimetro	16.5389
strumentale	6.9189	integrazione	7.0039	componente	16.0264
promiscuo	6.8682	co2e	6.9356	proprietario	14.8289
compensare	6.5809	finanziare	6.8713	capitale	14.7546
banking	6.026	stress	6.8447	societario	13.8689
area	5.8569	consulenza	6.3406	relazione	13.1863
responsible	5.5412	selezione	5.831	steinicke	12.7774
eq	5.5248	sgr	5.8241	materiale	12.44
raggiungere	5.5029	scope_3	5.7978	omissione	11.712
principles	5.3126	orientamento	5.3547	governo	11.6522
soddisfare	5.3126	obbligazione	4.9507	bank	11.0244
fornire	5.2718	climatico	4.8731	principio	10.0969

• *Banca Monte dei Paschi di Siena* (focuses on the role of emissions as a polluting factor and on the objectives to be achieved in environmental matters.

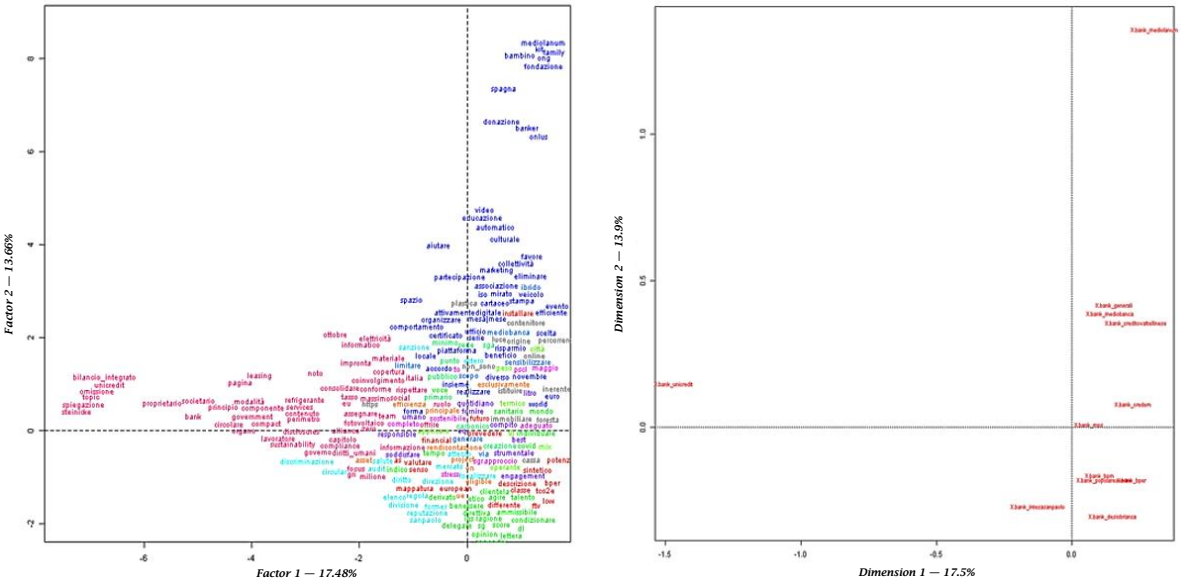
• *Banca Popolare di Sondrio* specifies the metrics used for reporting with the words "*tco2*" and "*scope-3*" and underlines the aspects of the European Directive on Integrated Reporting (IR).

• *Banca UniCredit* specifically reports on environmental "*principles*" and the reporting framework ("*gri*" or Global Reporting Initiative).

3.3. Lexical correspondence analysis (LCA)

The analysis of lexical correspondences (LCA) integrates with that of specificities and both identify the peculiar characteristics of a text.

Figure 4. Lexical correspondence analysis (LCA)



Source: Author's elaboration using Iramuteq.

LCA is an application to textual data of simple correspondence analysis (a multivariate statistical technique belonging to the group of factorial analyses[24]).

The Generali, Mediobanca, and Credito Valtellinese banks present a similar language that differs from UniCredit while Monte dei Paschi with its positioning close to the axes denotes a common language with the entire statistical collective.

[24] The large category of factorial analysis takes a very large data matrix and reduces it in dimensionality by projecting the large amount of data into a small subspace. This process is accomplished through the creation of factors — algebraically constructed variables - starting from the original data. The first two factors are taken, they are nothing more than the latent dimensions on which the data are projected, reproducing the maximum possible amount of variability that is contained in the original matrix. Factor analysis par excellence is the analysis of the principal components that works on quantitative variables; it creates the principal components (factors) — which are precisely a linear combination of the original variables; the use of the first two factors allows the representation in two dimensions. The analysis of lexical correspondences has the same objective as the analysis of the principal components: it synthesizes the content of many (qualitative) variables in a reduced subspace, then on the factors we project our information.

After examining the two overlapping factorial levels of the cross-analysis of the lexicon used by the banks, the following considerations can be made: the first dimension (first factor, 17.49% of total inertia represented), consists of the difference between lexicon which describes the practices of reducing environmental impacts, on the left, and the efficiency actions on the right.

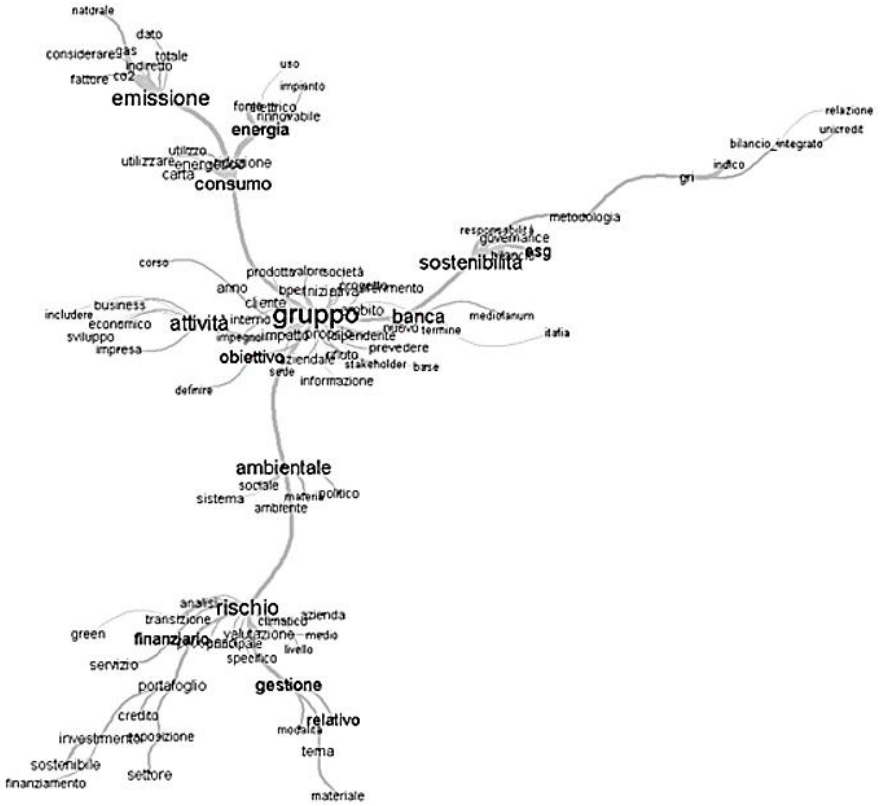
The second factorial dimension presents a polarization between the lexicon describing good social practices (typical of Banca Mediolanum) with the compliance characteristics deriving from international organizations and internal practices.

3.4. Similarities analysis

Similarity analysis identifies and represents co-occurrences within text segments (in Iramuteq they have 40 default units) allowing to study the construction of speech in a text.

Network analysis within textual analysis is a fairly recent approach in literature and economics: it moves the analysis from a frequency approach to a relational one where words have a relationship between them.

Figure 5. Similarities analysis



The most central words of the speech, written larger and located at the intersection points, are "*risk*", "*environmental*", "*consumption*", "*emission*", "*energy*", and "*sustainability*".

Excluding the graphic form "*group*", the study of the sequences clearly shows the concrete commitment to the reduction of emissions and the transition towards a low environmental impact economy.

It is believed that the construction of the discourse is linked to the transition from a carbon-intensive economy to one with low emissions.

In the context of the maximum three, there is also the word "*scenario*", it mainly refers to UniCredit, BPER, Credito Emiliano, and Mediobanca.

******001 *bank_unicredit *index_ftsemib *year_2021**

pubblicazione del primo rapporto tcfd 2020 separato nel mese di ottobre abbiamo iniziato ad applicare regolarmente la metodologia pacta per le banche per misurare l allineamento del nostro portafoglio creditizio a una serie di **scenari** climatici abbiamo aderito alla net zero

******004 *bank_bper *index_ftsemib *year_2021**

l approccio impiegato da bper per il calcolo del rischio di transizione sul portafoglio crediti è diretto a verificare il potenziale degli investimenti necessari ai settori a maggior rischio di transizione nell adeguarsi a due diversi **scenari** climatici presi a riferimento e può essere sintetizzato dall infografica seguente

******010 *bank_credem *index_ftseitaliamidcap *year_2021**

7 eventi climatici significativi del territorio italiano sopra riportati per ciascuna area geografica l analisi è stata condotta sulla base di due **scenari** futur

******010 *bank_credem *index_ftseitaliamidcap *year_2021**

scenario ottimistico di contrasto efficace al cambiamento_climatico e riduzione significativa delle emissioni di gas ad effetto serra nell atmosfera
scenario pessimistico comunemente associato all espressione business as usual o nessuna mitigazione in_cui la crescita delle emissioni continua ai ritmi attuali

3.5. Cluster analysis with Alceste (Reinert's method)

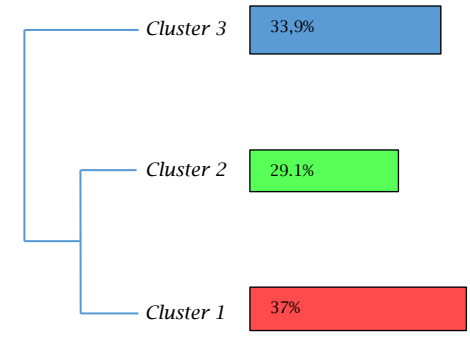
Cluster analysis, a method of unsupervised data mining, divides a set of statistical units so that the resulting subdivision has some properties considered desirable (Sokal & Rohlf, 1962), that is, to form groups of units so that the units that are assigned to the same group are homogeneous with respect to the characters considered, and each unit of the collection is contained in one and only one group (Gordon, 1999).

Cluster analysis is a set of relatively recent techniques that have interested statisticians since about 1960, also thanks to the parallel evolution of automatic calculation tools, which has made it possible to tackle computational complexities without difficulty.

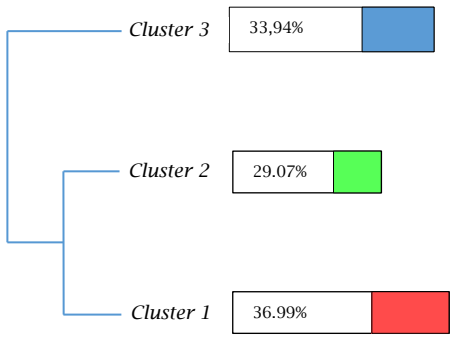
Iramuteq uses Reinert's method (Reinert, 1990) a descending agglomerative hierarchical method (hierarchical method) whose application creates three groupings.

Clustering has identified three lexical classes (Figure 6).

Figure 6. Dendrogram and cluster coverage



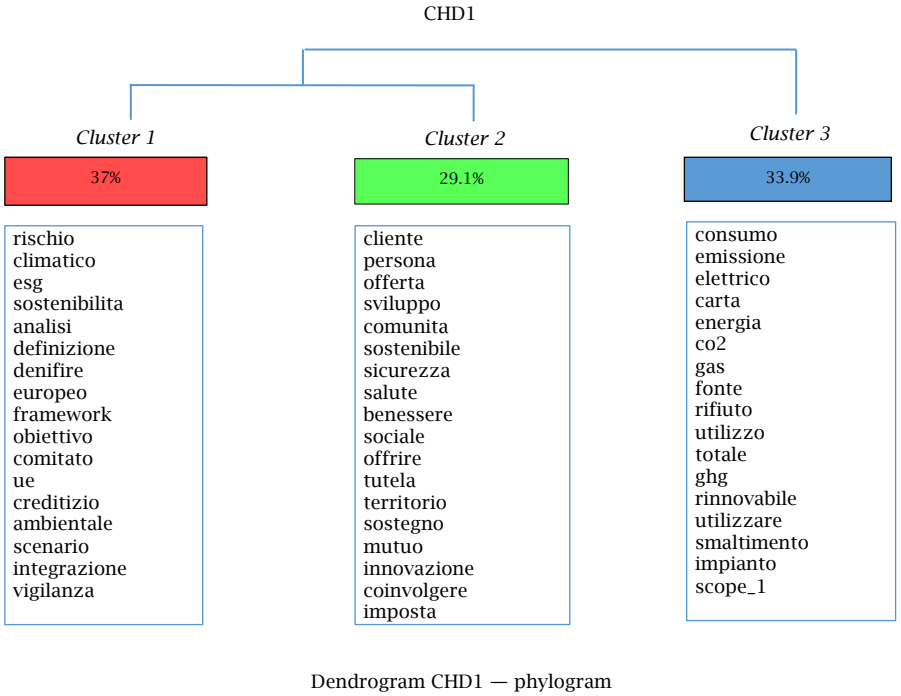
Dendrogram starting from CHD1



Dendrogram CHD1 — phylogram

The steps of the reclassification algorithm led to the identification of three final clusters represented by the dendrogram: Cluster 1 has 37%, 29.1% of classified segments falls into Cluster 2, and 33.94% of classified segments falls into Cluster 3.

Figure 7. Cluster analysis: Reinert's method



Alceste generates 2,398 segments of which 887 segments are included in Cluster 1, 697 in Cluster 2, and 814 in Cluster 3, out of a total of 2,398 segments.

Cluster 1 (red) refers to risks, climate change, and the frameworks used for reporting with the word "*scenario*" highlighting the need to provide future risk predictions.

Cluster 2 (green) specifies the centrality of the customer, the person, and the community, their well-being, the social aspects, and those connected to health, therefore it is substantially oriented to the social implications of the environmental aspects.

Cluster 3 (blue) is a universe of meaning linked to the practices of reducing emissions and environmental impact: "*consumption*", "*emission*", "*electricity*", "*waste*", "*ghg*", and "*scope_1*" testify to the actions of reduction of environmental risks.

3.6. Analysis with SAS Viya Text Mining

This part of the research is done with SAS Viya for Text Mining to locate and categorize data.

In its pipeline, there are five analysis nodes called: *concepts*, *parsing*[25], *sentiment*[26], *topics*[27] , and *categories* to allow you to identify the relevant textual data, and build "concept and categorization models" (SAS, 2019).

It was not possible to apply Machine Learning techniques due to the small size of the corpus and the same thing for sentiment analysis and topic detection.

Knowledge extraction with additional concepts is typical of the "*concepts*" node, it takes place with the creation of a set of concepts with the language interpretation and text interpretation (LITI) programming language.

The "*categories*" analysis node labels documents according to their content, by creating thought constructions that match research interests.

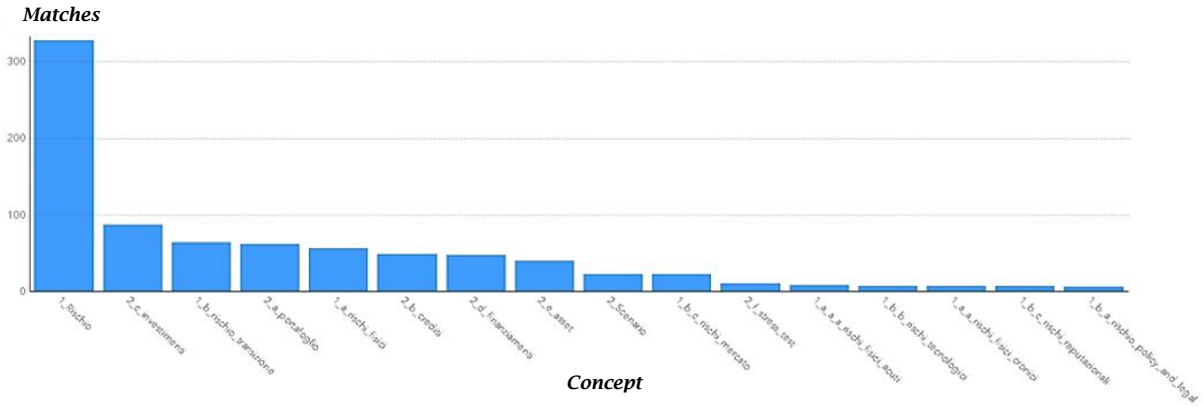
The programming codes in LITI language used in this research are provided in Appendix A.

The *concept* node analysis extracts specific information from the corpus on SAS Viya. Figure 9 expresses the number of matches per concept, i.e., how many times the concept appears in absolute terms in the text.

The word "*risk*" has an absolute frequency of 327 (three times higher than in 2020) while "*acute physical risks*" 8 and "*chronic physical risks*" 9, "*scenario*" has a frequency of 22 (lower than the previous year but concentrated in a few documents), "*policy and legal risk*" 6, "*market risks*" 22, "*technological risks*" and "*reputational risks*" 7 (see Figure 8). These words with their respective frequencies represent the typical focus on risk management of banks and the additional concepts outline a focus on ESG-offered products to customers like the previous year therefore the product offerings are sensitive to climate change criteria.

[25] It is the node that generates the term document matrix.
[26] The size of the file does not return a reliable sentiment analysis and so does the topic detection.
[27] The "*topics*" analysis node groups "similar documents into themes or topics", it is an automatic grouping of important terms that occur in documents and represent them; (first 5 terms that appear frequently in the topic).

Figure 8. The number of matches for the concept



The construction of the aforementioned concepts allows for contextual analysis and is useful for extracting information consistent with the research questions.

Figure 9. The number of documents by the concept

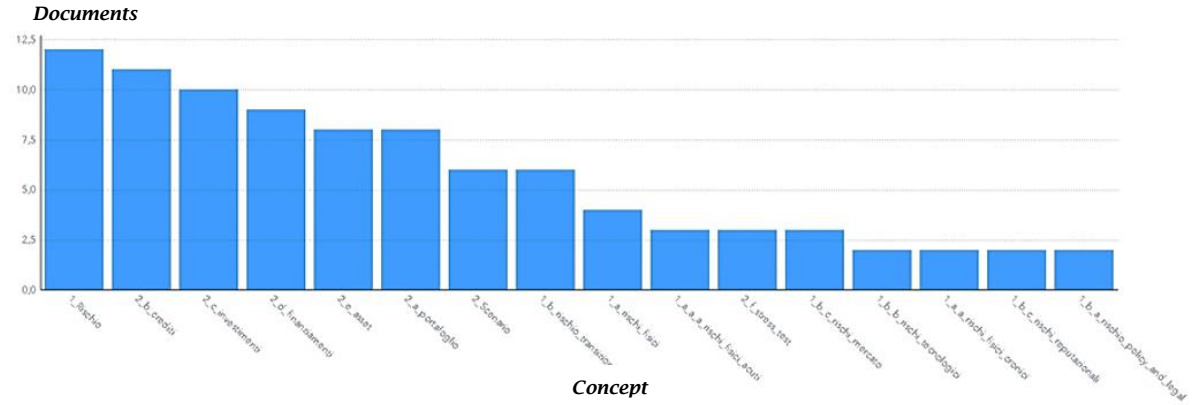


Figure 9 shows that the word "*scenario*" is present in 6 documents even if only in 4 documents, it is characterized by a full explanation of the analytical tool provided by the TCFD. "*Physical risks*" is present in 4 documents ("*acute physical risks*" 3 and "*chronic physical risks*" 2), and "*transition risk*" is present in 6 documents ("*policy and legal risk*" 2, "*market risks*" 3, "*technological risks*" 2, "*reputational risks*" 2). The word "*credits*" can be found in 11 documents, "*investments*" 10, "*loans*" 9, "*assets*" 8, and "*portfolio*" 8. It should be noted that "*scenario*" is a word present in BPER, BPM, Credito Emiliano, Mediobanca, and UniCredit.

Figure 10. The average number of matches per document

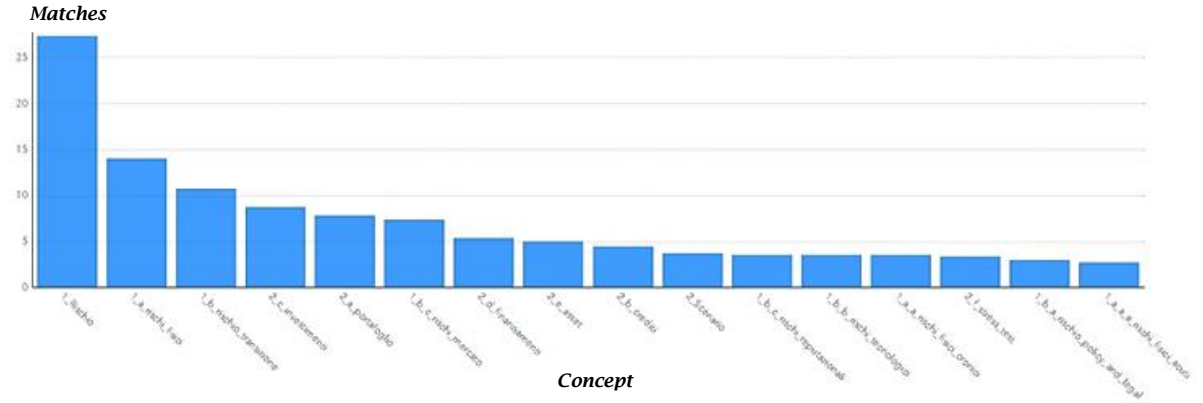


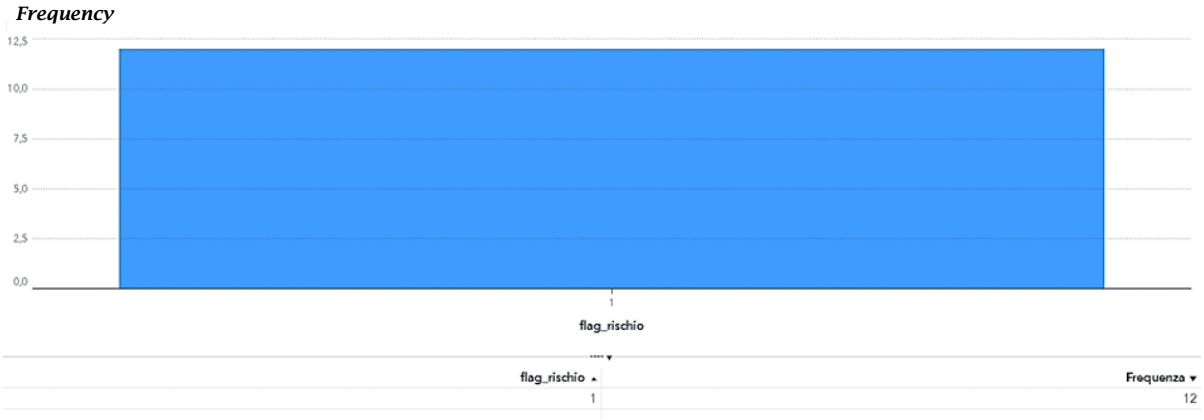
Figure 10 takes each frequency term and divides it by the number of documents: the average values using Tables 6 and 7. The word "*scenario*" only occurs in 6 documents and has an average value of 3.66. The words "*physical risks*" and "*transition*" are concentrated in a few documents having respective averages of 14 and 10.6.

The analysis node called *categories* labels documents according to their content, i.e., classifies documents by subject. A value of zero is interpreted

as the absence of the category and a value of 1 as the presence of the category (dummy variable), it also allows me to focus on new categories that correspond to my research interests.

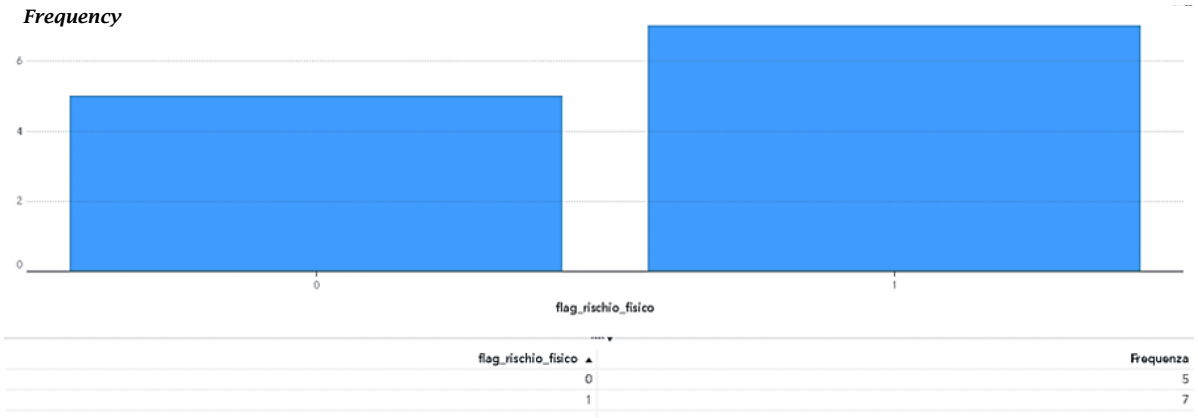
As seen in the node *concept*, the term "*risk*" is present in 12 documents and is evenly distributed (Figure 11).

Figure 11. The risk category



The category "*physical risk*" appears in 7 documents and is not present in 5 (Figure 12). It defines a lack of attention to this aspect of climate change, the banks concerned are BPER, BPM, Credito Emiliano, and Credito Valtellinese.

Figure 12. The physical risk category



The category "*acute physical risk*" appears in only 3 documents out of a total of 9 (Figure 13), while in the previous year, it was only present in the Credito Emiliano sustainability report (this is an improvement).

Figure 13. The acute physical risk category

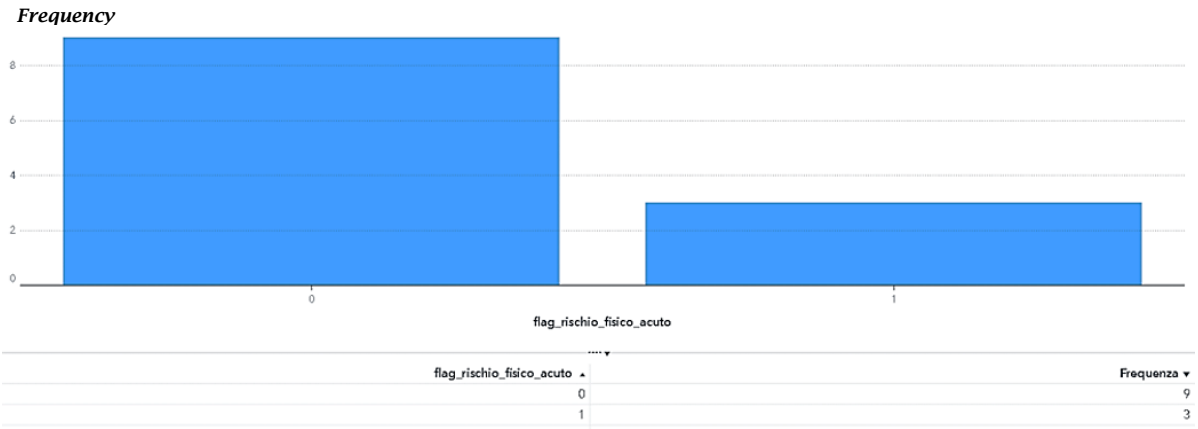
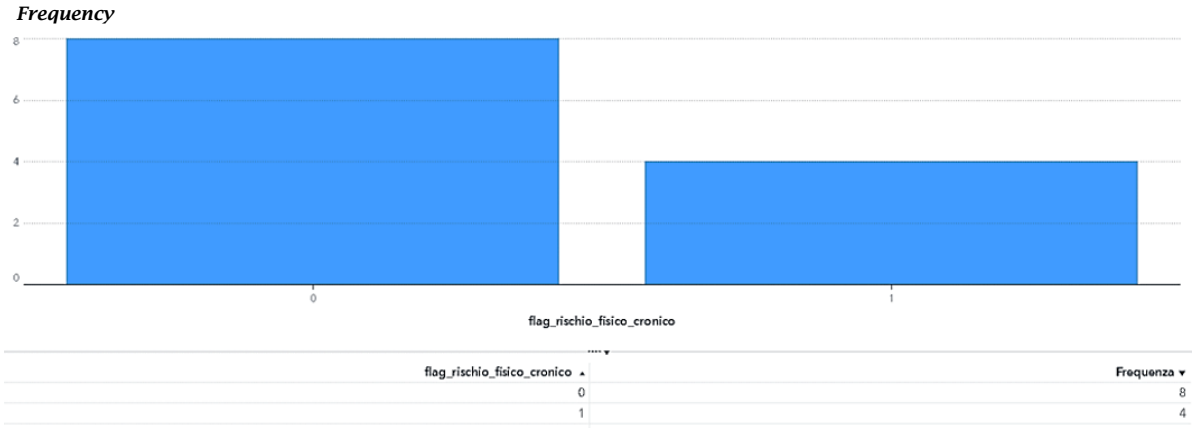
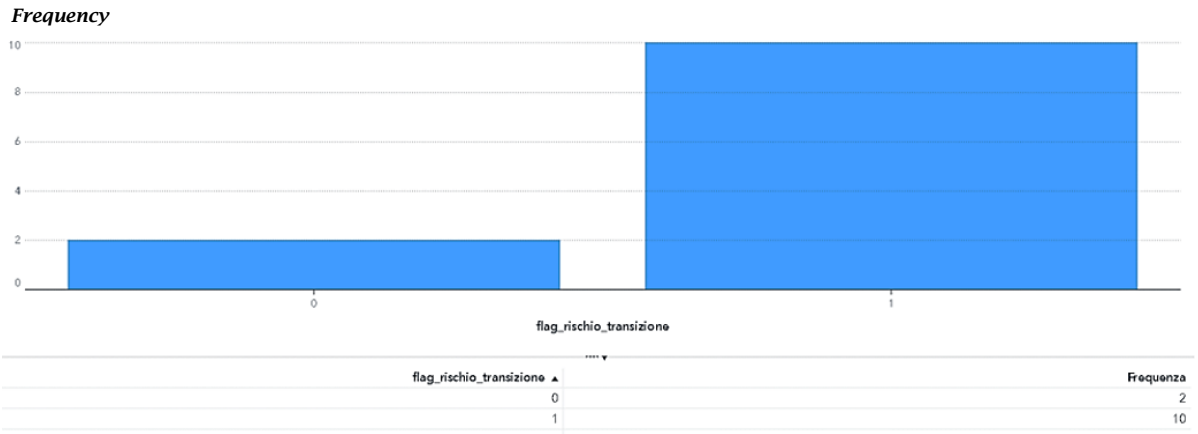


Figure 14. The chronic physical risk category



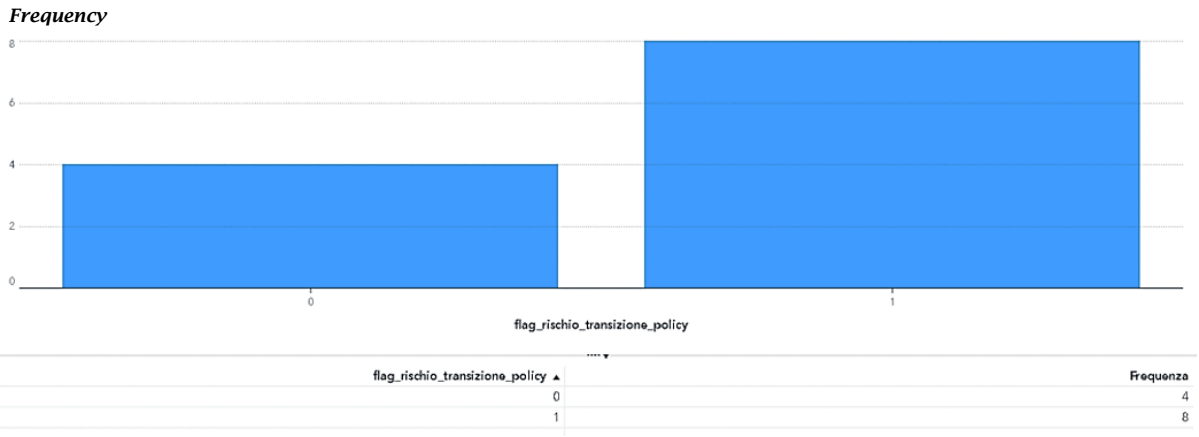
Similar considerations can be made for chronic physical risk because it is only encountered in 4 banks and not in 8 (Figure 14). Overall, it is evident that banks do not disclose these two important aspects.

Figure 15. The transition risk category



In Figure 15, "*transition risk*", which is identifiable in 10 out of 12 sustainability reports expresses the centrality of providing disclosure for the transition to a low-carbon economy.

Figure 16. The policy transition risk category



The breakdown of transition risk in its policy aspect shows a positive response in 66% of the documents (Figure 16).

The words "*transition*" and "*market*" conceptualized in a single variable are not encountered in 8 documents but only in 4 (Figure 17).

Figure 17. The market transition risk category

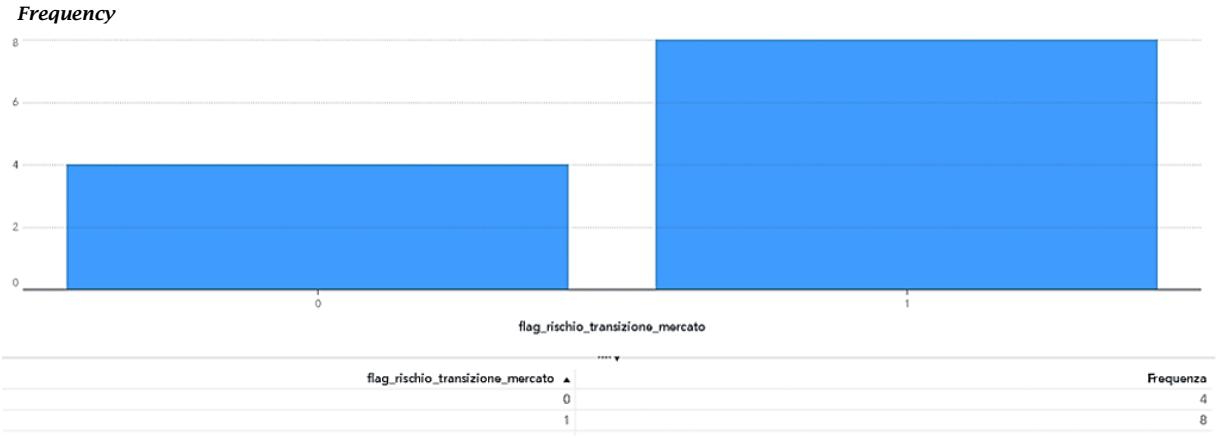
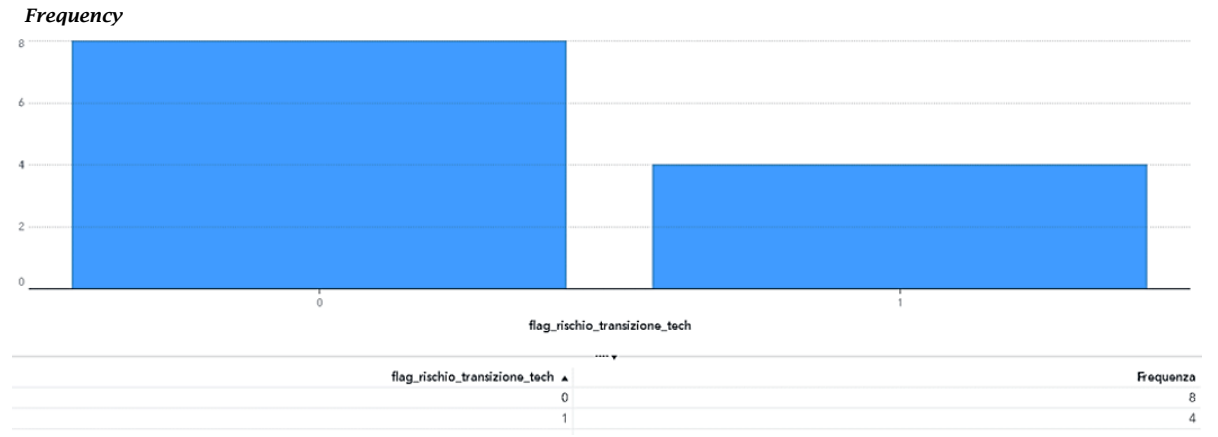


Figure 18. The technology transition risk category



The aspect of technology related to "*transition risk*" appears in only 4 documents and is absent in the other 8 (Figure 18).

"*Transition risk*" in its reputational articulation is not present in 10 documents (Figure 19).

Figure 19. The reputational transition risk category

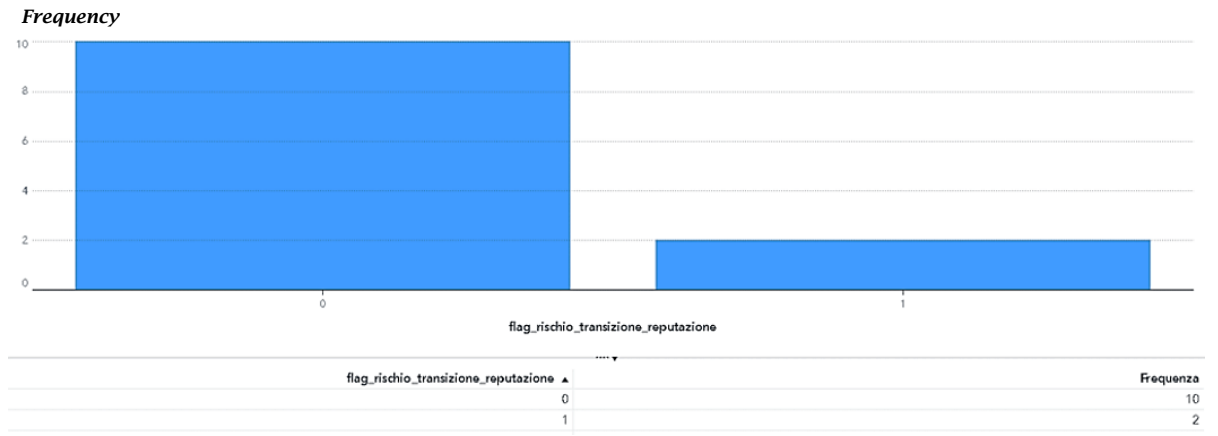
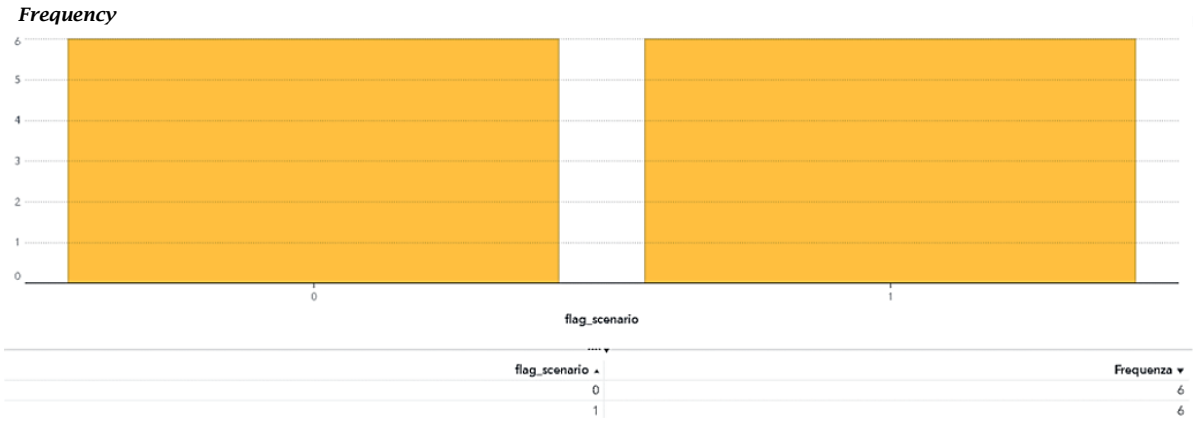


Figure 20. The scenario analysis category



The term "*scenario*" is present in 6 banks and compared to the previous year the lack of interest in this forecasting tool is reduced.

There are 4 banks that carry out scenario analyses and are compliant with the TCFD: BPER, Credito Emiliano, Mediobanca, and UniCredit.

Banca BPER signed the framework in 2021, Banca Credito Emiliano has undertaken scenario analysis (business as usual) and provided Scope 3 information in line with the TCFD metrics, disclosures on physical and transition risks by integrating climate aspects into corporate strategy and governance, Banca UniCredit paid close attention to climate change using scenario forecasts and in October of the year 2021 produced a separate report compliant with the TCFD framework, Mediobanca gave ample disclosure of climate risks, of Scope 1 and 2 emissions (market-based and location-based), also of Scope 3 and identifies climatic risks with the prescriptions of the TCFD.

4. RESULTS AND DISCUSSION

The key findings are still the inadequacy and incompleteness of sustainability reports (Aversa, 2023a) like the previous year of analysis, even if a clear improvement was made compared to the previous year.

The use of pillars and TCFD recommendations by BPER, Credito Emiliano, Mediobanca, and UniCredit is a substantial step forward to disclose forward-looking and firm-specific information for investors, stakeholders, and regulators.

Despite this significant improvement, over half of the banks (66%) provide incomplete information that leads to incomparability and reduced transparency of the ESG disclosures.

The banks contribute very moderately to reducing the sustainability data gap with disclosures that do not have a unique and clearly used framework.

In banks, it persists to pronounced opacity in the assessment of physical risks compared to transition risks which requires an improvement of the risk management functions. Concerning alignment to the pillars and recommendations of the TCFD only 4 sustainability reports were made with these criteria (banks provided ESG information without prioritizing climate change risks).

5. CONCLUSION

In conclusion, reducing the fragmentation of information and rapid action on shared forms of disclosures need an integration path with the TCFD pillars and recommendations; regulators and governments must enforce in law TCFD disclosures making them mandatory. Disclosing climate-related financial information on a mandatory basis allows for increased transparency and more comparability on corporate sustainability reports helps investors and businesses to better understand the financial impact of their exposure, and provides clearer disclosures to actual and potential investors, lenders, and regulators. In this way, an economic system with low carbon emissions and high energy efficiency is achieved, reducing uncertainty in the markets and reducing the potential for financial instability.

With regards to disclosure, the harmonization of sustainability reporting standards through the Corporate Sustainability Reporting Directive[28] (CSRD), Sustainable Finance Disclosure Regulation (SFDR, 2019/2088)[29], and European Taxonomy Directive (2020/852), "should reduce the sustainability data gap within the European Union" (Banca d'Italia, 2021) and improve quality-quantitative aspects in the future.

It is worth noting that there is a lack of real coordination with the provisions of the TCFD whose references to the framework seem to be purely declamatory and without real consequential actions.

The Implementing Technical Standard (defined by the EBA in January 2022[30]) with templates for a) qualitative disclosure of climate risks, b) quantitative disclosure of transition risk, c) quantitative disclosure of physical risks related to climate change, and d) the use of certain KPIs, the Green Asset Ratio (GRA) and the Banking Book Taxonomy Alignment Ratio (BTAR) could significantly improve things.

In this regard, central banks need to be organized with staff capable of web scraping and web crawling the necessary information from websites, precisely taking into account the web disclosing obligations of Articles 7 and 9 of the SFDR.

[28] CSRD supersedes the Non-Financial Reporting Directive (NFRD, 2014/95/EU) transposed into Italian law by legislative decree 254/2016.
[29] In the future the requirements on climate disclosures with the International Sustainability Standard Board (ISSB) will play a new role.
[30] ITS (P3 ESG ITS).

About the sustainability data gap, textual data are unstructured data of human origin, they do not conform to predefined patterns and are rather complex to analyze; the enormous production of textual data introduces computational challenges for which gathering qualitative and quantitative data (as suggested by NGFS) can take place with machine learning tools combining qualitative and quantitative variables.

Classical (inferential) statistics extend what we observe on small samples to a larger population and this is done through fairly simple models such as the linear one, while with very large datasets and more powerful models such as machine learning, more sophisticated relationships can be sought.

In addition, since there is no sampling procedure when dealing with large amounts of data (big data), there is consequently no problem with dealing with sampling error (in machine learning, the measurement of error is aimed at evaluating distortion and reliability of the model together with the data quality analysis.

It can be said that big data differs from traditional data, generally based on samples, they must manage sample variability and therefore give rise to the inductive problem of passing from the sample to the population through probabilistic tools (moreover, the sample can provide information that is not necessarily similar to those of the population).

Inference combines probability and inductive process thereby simple hypotheses can be made on the sample, with simple models (typically the linear one), because it is characterized by a limited + number of observations; when instead the dataset is large, we do not have the problems of sample variability.

In big data, as variables grow, so does the number of spurious correlations (correlations between variables that do not make sense) that can lead to false scientific discoveries and erroneous statistical inferences.

At the same time, if we have a large sample size, it makes no sense to carry out hypothesis tests; they lose their significance since with many units they are always significant: in fact, all hypothesis tests have the sample size as their denominator.

Furthermore, multiple tests lose even more meaning, such as in the stepwise procedures where variables are added and removed from the model; this is an incorrect procedure as each time the level of significance should be re-established as the test progresses (and taken into account the dependence of the tests).

It should be said that for classical statistical methods, analyzing large databases (e.g., with a million observations) is possible through unit sampling, while for machine learning methods, a large number of observations is a necessity.

Conversely, if there is a large number of variables (e.g., thousands of variables) the classical method is difficult to apply while the machine learning methods could be applied more effectively and successfully even if it is often preferable to first go through a reduction of the input variables.

Gathering qualitative and quantitative data allows us to obtain complementary information that cannot be provided by quantitative data, and can be used to improve the structure and partition of datasets.

We add some considerations about uncertainty and deep uncertainty. Traditional uncertainty can be addressed with probabilistic models while deep uncertainty is characterized by limited available information, insufficient historical data, and variables interconnected in complex and not yet understood ways.

To the best of our knowledge, central banks used a simple concept of uncertainty in climate change, namely the concept of uncertainty related to. The exponent of the Chicago School distinguishes between risk and uncertainty in these terms: one speaks of risk when it is possible to calculate the probability of an event, and one speaks of uncertainty when it is not possible to calculate the probability of the event.

Climate phenomena, on the other hand, exhibit characteristics of non-linearity for which the concept of deep uncertainty would seem to be better suited, since determinist and probabilistic methods are insufficient for analysis and it seems necessary to treat uncertainty in models with non-additive/ additive (probabilities (probabilities whose sum may be less than or greater than 1).

Certainly, a decision by a central bank must have a strong theoretical justification, but we might think that the use of more sophisticated (and in my opinion more suitable) concepts of uncertainty can be justified in terms of policy and pending a completed theoretical elaboration[31].

In addition to the topics of disclosure, inference, and uncertainty, it should be added that the "gathering of qualitative and quantitative data" to combat climate change is required among others by BIS and NGFS; some publications call for new methods to process the data, e.g., Banca d'Italia proposes to use machine learning but does not go into detail on the techniques to be used.

To address the shortcomings of sustainability data gaps central banks, have to deal with unstructured data (video, audio, and text) via text mining and the processing of this data is not trivial. Indeed, the transformation of words into numbers[32] is neither simple nor unambiguous, it would lead central banks to use much more powerful analysis tools such as deep learning (neural networks) and transformers.

For scenario analyses, the same methodological considerations as above are to be applied, and it can be emphasized, however, that climate change has been addressed almost exclusively in terms of economic policy: carbon tax and cap-and-trade ("hoping that market forces alone will be sufficient to achieve the policy objective") while the current debate draws attention to the immediate urgency for central banks to help keep global warming below 2°C by scaling up green finance and preserving financial stability (D'Orazio & Dirks, 2022).

A personal proposal could be to differentiate the carbon tax by means of taxation instruments that act not only on direct but also on indirect

[31] Of course, the question remains open as to how reliably the monetary policy authority views the model of the economy it is using (the author is thinking of the work of Hansen and Sargent, 2022 and then Hansen's for more popular presentations).

[32] The techniques can be summarized in: semantic representation and ASCII, one-hot-representation, vector space models.

taxation; there is a huge difference in directly hitting turnover instead of indirectly hitting consumption i.e., value-added-tax (VAT); combinations of the two instruments may be necessary and various kinds of simulations are required, for which agent-based models seem more suitable. We think this distinction should also be introduced into climate change projection models and software, in the author's opinion, some of them seem overly sensitive to direct taxation assumptions.

The paper will be very important for future research that will assess climate change risks in the banking sector and the relative disclosures via text mining while a limit of the work is due to the small size of the corpora that does not allow the use of machine learning methods, especially LLMs which present themselves as an alternative and more complete than topic modeling techniques.

REFERENCES

1. Angelico, C., Faiella, I., & Michelangeli, V. (2022). Il rischio climatico per le banche italiane: Un aggiornamento sulla base di un indagine campionaria. *Note di Stabilità Finanziaria e Vigilanza, 29*. https://www.bancaditalia.it/pubblicazioni/note-stabilita/2022-0029/index.html
2. Ansar, A., Caldecott, B., & Tilbury, J. (2013). *Stranded asset and the fossil fuel divestment campaign: What does divestment mean for the valuation of fossil fuel assets?* (Working paper). Smith School of Enteprise and the Environment, University of Oxford. https://www.smithschool.ox.ac.uk/sites/default/files/2022-03/SAP-divestment-report-final.pdf
3. Aversa, D. (2023a). Climate change and climate-related financial disclosures in the banking sector. *Risk Governance and Control: Financial Markets and Institutions, 13*(1), 70–94. https://doi.org/10.22495/rgcv13i1p6
4. Aversa, D. (2023b). Scenario analysis and climate change: A literature review via text analytics. *British Food Journal, 126*(1), 271-289. https://doi.org/10.1108/BFJ-08-2022-0691
5. Aversa, D., & Cincinelli, P. (2019). The contribution of Italian banking system to the Italian economic and social growth. In D. Aversa (Ed.), *Politics of banking: Regulatory capture, thin political markets, ideology and beyond* (pp. 53–84). Aracne.
6. Aversa, D., Adamashvili, N., Fiore, M., & Spada, A. (2022). Scoping review (SR) via text data mining on water scarcity and climate change. *Sustainability, 15*(1), Article 70. https://doi.org/10.3390/su15010070
7. Banca d'Italia. (2021). *Aspettative di vigilanza sui rischi climatici e ambientali.* https://www.bancaditalia.it/focus/finanza-sostenibile/vigilanza-bancaria/Aspettative_di_vigilanza_BI_su_ESG.pdf
8. Barnett, M., Brock, W., & Hansen, L. P. (2022). Climate change uncertainty spillover in the macroeconomy. *NBER Macroeconomics Annual, 36*(1), 253–320. https://doi.org/10.1086/718668
9. Basel Committee on Banking Supervision (BCBS). (2020, April). *Climate-related financial risks: A survey on current initiatives.* Bank for International Settlements (BIS). https://www.bis.org/bcbs/publ/d502.htm
10. Basel Committee on Banking Supervision (BCBS). (2021, April). *Climate-related risk drivers and their transmission channels.* Bank for International Settlements (BIS). https://www.bis.org/bcbs/publ/d517.htm
11. Battiston, S., Dafermos, Y., & Monasterolo, I. (2021). Climate risks and financial stability. *Journal of Financial Stability, 54*, Article 100867. https://doi.org/10.1016/j.jfs.2021.100867
12. Bernardini, E., Faiella, I., Lavecchia, L., Mistretta, A., & Natoli, F. (2021). *Banche centrali, rischi climatici e finanza sostenibile* (Questioni di Economia e Finanza Occasional Paper No. 608). Banca d'Italia. https://www.bancaditalia.it/pubblicazioni/qef/2021-0608/index.html
13. Bholat, D. (2015). Big Data and central banks. *Big Data & Society, 2*(1). https://doi.org/10.1177/2053951715579469
14. Camargo, B. V., & Justo, A. M. (2016). IRaMuTeQ: R interface for multidimensional analysis of texts and questionnaires. *Um Software Gratuito para Análise de Dados Textuais.Temas em Psicologia, 21*(2), 513–518. http://www.iramuteq.org/documentation/fichiers/IRaMuTeQ%20Tutorial%20translated%20to%20English_17.03.2016.pdf
15. Carney, M. (2013, August 28). *Crossing the threshold to recovery* [Speech]. The Bank of England. https://www.bankofengland.co.uk/-/media/boe/files/speech/2013/crossing-the-threshold-to-recovery-transcript.pdf
16. Cerreia-Vioglio, S., Hansen, L. P., Maccheroni, F., & Marinacci, M. (2021). *Making decisions under model misspecification* (University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-103). https://doi.org/10.2139/ssrn.3666424
17. Chenet, H., Ryan-Collins, J., & van Lerven, F. (2021). Finance, climate-change and radical uncertainty: Towards a precautionary approach to financial policy. *Ecological Economics, 183*, Article 106957. https://doi.org/10.1016/j.ecolecon.2021.106957
18. Coueré, B. (2018, November 8). *Monetary policy and climate change* [Speech]. European Central Bank. https://www.ecb.europa.eu/press/key/date/2018/html/ecb.sp181108.en.html
19. D'Orazio, P. & Dirks, M. W. (2022). Exploring the effects of climate-related financial policies on carbon emissions in G20 countries: A panel quantile regression approach. *Environmental Science and Pollution Research, 29*(5), 7678–7702. https://doi.org/10.1007/s11356-021-15655-y
20. D'Orazio, P., Popoyan, L., & Valente, M. (2023). Green central banking in times of climate uncertaitny. In S. Alam & S. Sosnovskikh (Eds.), *Environmental finance and green banking: Contemporary and emerging issues.* Routledge. https://doi.org/10.4324/9781003206194-5
21. Ebner, A. (2018). The financial stability aspects of the EU-wide stress test. *Journal of Financial Regulation, 4*(2), 326–336. https://doi.org/10.1093/jfr/fjy009
22. Eccles, R. G., & Krzus, M. P. (2018). Implementing the task force on climate-related financial disclosures recommendations: An assessment of corporate readiness. *Schmalenbach Business Review, 71*(2), 287–293. https://doi.org/10.1007/s41464-018-0060-4
23. Emambakhsh, T., Giuzio, M., Mingarelli, L., Salakhova, D., & Spaggiari, M. (2022, May). Climate-related risks to financial stability. In *Financial Stability Review*. European Central Bank. https://www.ecb.europa.eu/pub/financial-stability/fsr/special/html/ecb.fsrart202205_01~9d4ae00a92.en.html

24. European Central Bank (ECB). (2021, July). *Climate-related risk and financial stability*. https://www.ecb.europa.eu/pub/pdf/other/ecb.climateriskfinancialstability202107~87822fae81.en.pdf
25. Feridun, M., & Güngör, H. (2020). Climate-related prudential risks in the banking sector: A review of the emerging regulatory and supervisory practices. *Sustainability, 12*(13), Article 5325. https://doi.org/10.3390/su12135325
26. GARP. (2022). *SCR certificate sustainability and climate risk exam*. GARP.
27. Giuliano, L. (2004). *L'analisi automatica dei dati testuali: Software e istruzioni per l'uso*. LED Edizioni Universitarie.
28. Gordon, A. D. (1999). Cluster validation. In C. Hayashi, K. Yajima, H. H. Bock, N. Ohsumi, Y. Tanaka, & Y. Baba (Eds.), *Proceedings of the Fifth Conference of the International Federation of Classification Societies "Data science, classification, and related methods"* (pp. 22–39). Springer, Tokyo. https://doi.org/10.1007/978-4-431-65950-1_2
29. Hansen, J., Sato, M., & Ruedy, R. (2012). Perception of climate change. *Proceedings of the National Academy of Sciences (PNAS), 109*(37), E2415–E2423. https://doi.org/10.1073/pnas.1205276109
30. Hansen, L. P., & Miao, J. (2018). Aversion to ambiguity and model misspecification in dynamic stochastic environments. *Proceedings of the National Academy of Sciences (PNAS), 115*(37), 9163–9168. https://doi.org/10.1073/pnas.1811243115
31. Hansen, L. P., & Sargent, T. J. (2022). Structured ambiguity and model misspecification. *Journal of Economic Theory, 199*, Article 105165. https://doi.org/10.1016/j.jet.2020.105165
32. Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *The Quarterly Journal of Economics, 133*(2), 801–870. https://doi.org/10.1093/qje/qjx045
33. Houghton, J. T., Jenkins, G. J., & Ephraums. J. J. (1990). *Climate change: The IPCC scientific assessment*. Cambridge University Press. https://archive.ipcc.ch/ipccreports/far/wg_I/ipcc_far_wg_I_full_report.pdf
34. Hwang, H., An, S., Lee, E., Han, S., & Lee, C.-h. (2021). Cross-societal analysis of climate change awareness and its relation to SDG 13: A knowledge synthesis from text mining. *Sustainability, 13*(10), Article 5596. https://doi.org/10.3390/su13105596
35. Intergovernmental Panel on Climate Change (IPCC). (1995). *Climate change, economic and social dimensions of climate change*.
36. Intergovernmental Panel on Climate Change (IPCC). (2001). *Climate change, the scientific basis*.
37. International Sustainability Standards Board (ISSB). (2022). *Exposure draft IFRS® sustainability disclosure standard: [Draft] IFRS S2 climate-related disclosures*. International Financial Reporting Standards (IFRS) Foundation.
38. Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2020). *Making text count: Economic forecasting using newspaper text* (Bank of England Working Paper No. 865). https://doi.org/10.2139/ssrn.3610770
39. Kandlikar, M., Risbey, J., & Dessai, S. (2005). Representing and communicating deep uncertainty in climate-change assessments. *Comptes Rendus. Geoscience, 337*(4), 443–455. https://doi.org/10.1016/j.crte.2004.10.010
40. McPhail, C., Maier, H. R., Westra, S., Kwakkel, J. H., & van der Linden, L. (2020). Impact of scenario selection on robustness. *Water Resources Research, 56*(9), Article e2019WR026515. https://doi.org/10.1029/2019WR026515
41. Minenna, M., & Aversa, D. (2019). A revised European stability mechanism to realize risk sharing on public debts at market conditions and realign economic cycles in the Euro area. *Economic Notes, 48*(1), Article 12118. https://doi.org/10.1111/ecno.12118
42. Moreno, A.-I., & Caminero, T. (2022). Application of text mining to the analysis of climate-related disclosures. *International Review of Financial Analysis, 83*, Article 102307. https://doi.org/10.1016/j.irfa.2022.102307
43. Network for Greening the Financial System (NGFS). (2020, June). *NGFS climate scenarios for central banks and supervisors*. https://www.ngfs.net/en/ngfs-climate-scenarios-central-banks-and-supervisors
44. Nyambuu, U., & Semmler, W. (2023). *Sustainable macroeconomics, climate risks and energy transitions: Dynamic modeling, empirics, and policies*. Springer, Cham. https://doi.org/10.1007/978-3-031-27982-9
45. Official Monetary and Financial Institutions Forum (OMFIF). (2020). *Tackling climate change: The role of banking regulation and supervision*. https://www.omfif.org/tacklingclimatechange/
46. Parry, M., Canziani, O., Palutikof, J., van der Linden, P., & Hanson, C. (2007). *Climate change 2007: Impacts, adaptation and vulnerability*. Intergovernmental Panel on Climate Change (IPCC). https://www.ipcc.ch/site/assets/uploads/2018/03/ar4_wg2_full_report.pdf
47. Ramani, V. (2020). *Addressing climate as a systemic risk: A call to action for financial regulators*. Ceres. https://www.ceres.org/sites/default/files/reports/2020-06/Financial%20Regulators%20FULL%20FINAL.pdf
48. Reinert, M. (1990). Alceste une méthodologie d'analyse des données textuelles et une application: Aurelia De Gerard De Nerval. *Bulletin de Méthodologie Sociologique, 26*(1), 24–54. https://doi.org/10.1177/075910639002600103
49. SAS. (2019). *Exploring SAS Viya: Visual analytics, statistics, and investigations*. https://support.sas.com/content/dam/SAS/support/en/books/free-books/exploring-sas-viya-va-statistics-investigations.pdf
50. Schnabel, I. (2020, September 28). *When the markets fail: The need for collective action in tackling climate change* [Speech]. European Central Bank. https://www.ecb.europa.eu/press/key/date/2020/html/ecb.sp200928_1~268b0b672f.en.html
51. Siderius, K. (2023). An unexpected climate activist: Central banks and the politics of the climate-neutral economy. *Journal of European Public Policy, 30*(8), 1588–1608. https://doi.org/10.1080/13501763.2022.2093948
52. Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon, 11*(2), 33–40. https://doi.org/10.2307/1217208
53. Souza, M. A. R. d., Wall, M. L., de Morais Chaves Thuler, A. C., Lowen, I. M. V., & Peres, A. M. (2018). O uso do software IRAMUTEQ na análise de dados em pesquisas qualitativas [The use of IRAMUTEQ software for data analysis in qualitative research]. *Revista Da Escola De Enfermagem Da USP, 52*, Article e03353. https://doi.org/10.1590/S1980-220X2017015003353
54. Stern, N., Stiglitz, J., & Taylor, C. (2022). The economics of immense risk, urgent action and radical change: Towards new approaches to the economics of climate change. *The Journal of Economic Methodology, 29*(3), 181–216. https://doi.org/10.1080/1350178X.2022.2040740

55. Svartzman, R., Bolton, P., Despres, M., Pereira Da Silva, L. A., & Samama, F. (2021). Central banks, financial stability and policy coordination in the age of climate uncertainty: A three-layered analytical and operational framework. *Climate Policy, 21*(4), 563–580. https://doi.org/10.1080/14693062.2020.1862743
56. Task Force on Climate-related Financial Disclosures (TCFD). (2017a). *Recommendations of the Task Force on Climate-related Financial Disclosure* (Final report). https://assets.bbhub.io/company/sites/60/2020/10/FINAL-2017-TCFD-Report-11052018.pdf
57. Task Force on Climate-related Financial Disclosures (TCFD). (2017b). *The use of scenario analysis in disclosure of climate-related risks and opportunities* (Technical supplement). https://assets.bbhub.io/company/sites/60/2020/10/FINAL-TCFD-Technical-Supplement-062917.pdf
58. Task Force on Climate-related Financial Disclosures (TCFD). (2019). 2019 Status report. https://assets.bbhub.io/company/sites/60/2020/10/2019-TCFD-Status-Report-FINAL-053119.pdf
59. Task Force on Climate-related Financial Disclosures (TCFD). (2020a, October) *Forward-looking financial sector metrics consultation*. https://assets.bbhub.io/company/sites/60/2021/07/TCFD_Consultation_ForwardLooking_Final.pdf
60. Task Force on Climate-related Financial Disclosures (TCFD). (2020b). *2020 Status report*. https://www.fsb.org/wp-content/uploads/P291020-1.pdf
61. Task Force on Climate-related Financial Disclosures (TCFD). (2021). *2021 Status report*. https://www.fsb.org/wp-content/uploads/P141021-1.pdf
62. Task Force on Climate-related Financial Disclosures (TCFD). (2022). *2022 Status report*. https://assets.bbhub.io/company/sites/60/2022/10/2022-TCFD-Status-Report.pdf
63. The Bank of England. (2015). *One bank research agenda* (Discussion paper). https://shorturl.at/lyBPY
64. The Bank of England. (2021). *Remit for monetary policy committee* (Technical report).
65. The Bank of England. (2022). *Results of the 2021 climate biennial exploratory scenario (CBES)*. https://www.bankofengland.co.uk/stress-testing/2022/results-of-the-2021-climate-biennial-exploratory-scenario
66. Tipaldo, G. (2014). *L'analisi del contenuto e i mass media: Oggetti, metodi e strumenti*. il Mulino.
67. Tol, R. S. J. (2023). *Climate economics: Economic analysis of climate, climate change and climate policy* (3rd ed.). Edward Elgar.
68. U.S. Governing Publishing Office Washington. (2021). *Addressing climate as a systemic risk*. U.S. Government Publishing Office.
69. United Nations Climate Change. (n.d.). *The Paris Agreement. What is the Paris Agreement*? https://unfccc.int/process-and-meetings/the-paris-agreement
70. Walker, W. E., Lempert, R. J., & Kwakkel, J. H. (2013). Deep uncertainty. In S. I. Gass & M.C. Fu (Eds.), *Encyclopedia of operations research and management science* (pp. 395–402). Springer. https://doi.org/10.1007/978-1-4419-1153-7_1140
71. Zhai, P. (2013). *Climate change, the physical science basis*. Intergovernmental Panel on Climate Change (IPCC). https://apps.ipcc.ch/outreach/documents/360/1477282598.pdf

APPENDIX

A.1. Concepts

Risk
CLASSIFIER: risk
CLASSIFIER: risks

Physical risk
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "physical")
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "physical")
CONCEPT_RULE: (DIST_20, "_c{damage}", "physical")
CONCEPT_RULE: (DIST_20, "_c{damage}", "physical")

Chronic physical risk
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "chronic")
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "chronic")
CONCEPT_RULE: (DIST_20, "_c{1_a_Rischi_Fisici}", "chronic")
CONCEPT_RULE: (DIST_20, "_c{1_a_Rischi_Fisici}", "chronic")

Acute physical risk
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "acute")
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "treble")
CONCEPT_RULE: (DIST_20, "_c{1_a_Rischi_Fisici}", "acute")
CONCEPT_RULE: (DIST_20, "_c{1_a_Rischi_Fisici}", "treble")

Transition risk
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "transition")

Policy and legal risk
CONCEPT_RULE: (DIST_20, "_c{1_b_Rischi_transizione}", "normative")
CONCEPT_RULE: (DIST_20, "_c{1_b_Rischi_transizione}", "policy")

Technology risk
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "technological")
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "technological")
CONCEPT_RULE: (DIST_20, "_c{technological}", "risk")
CONCEPT_RULE: (DIST_20, "_c{technology}", "risk")

Market risk
CONCEPT_RULE: (DIST_20, "_c{1_Rischio}", "market")
CONCEPT_RULE: (DIST_20, "_c{market}", "risk")

Reputation risk
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "reputational")
CONCEPT_RULE: (DIST_20, "_c{1_Risk}", "reputational")

Mitigation
CLASSIFIER: mitigation
CLASSIFIER: mitigate

Adaptation
CLASSIFIER: adaptation

Resilience
CLASSIFIER: resilience
CLASSIFIER: resilient
CLASSIFIER: resilient
CLASSIFIER: resilience

Covid
CLASSIFIER: covid
CLASSIFIER: Covid
CLASSIFIER: COVID

Scenario
CLASSIFIER: scenario

A.2. Categories

Environment_Flag
(OR, "Environment", "environment", "Environmental", "environmental", "Environment", "environment", "Environmental", "environmental")

Flag_Clima
(OR, "climate", "Climate", "climate", "Climate", "climate", "Climate")

Flag_Covid
(OR, "covid", "Covid", "COVID")

Flag_Emission
(OR, "Emission", "emission", "Emissions", "emissions")

Flag_Mitigation_Adaptation
(OR, "mitigation", "mitigate", "adapt", "adapt")

Flag_Resilience
(OR, "resilience", "resilient", "resilience", "resilient")

Flag_Risk
(OR, "risk", "risks", "risk")

Flag_Physical_Risk
(AND, (OR, "damage", "damage", "risk", "risks"), (OR, "physical", "physical"))

Acute_Physical_Risk_Flag
(AND, (OR, "harm", "harm", "risk", "risk"), (OR, "acute", "acute"))

Flag_Risk_Physical_Chronic
(AND, (OR, "harm", "harm", "risk", "risk"), (OR, "chronic", "chronic"))

Transition_Risk_Flag
(AND, (OR, "risk", "risks"), (OR, "transition"))

Market_Transition_Risk_Flag
(OR, (AND, (AND, (OR, "risk", "risks"), "transition"), "markets"), (AND, (AND, (OR, "risk", "risks"), "transition"), "market"))

Policy_Risk_Transition_Flag
(OR, (AND, (AND, (OR, "risk", "risks"), "transition"), "policy"), (AND, (AND, (OR, "risk", "risks"), "transition"), "normative"), (AND, (AND, (OR," risk", "risks"), "transition"), "legal"), (AND, (AND, (OR, "risk", "risks"), "transition"), "legal"))

Flag_Risk_Transition_Reputation
(OR, (AND, (AND, (OR, "risk", "risks"), "transition"), "reputational"), (AND, (AND, (OR, "risk", "risks"), "transition"), "reputational"))

Tech_Transition_Risk_Flag
(OR, (AND, (AND, (OR, "risk", "risks "), "transition"), "technological"), (AND, (AND, (OR, "risk", "risks "), "transition"), "technological"))

Flag_Scenario
(OR, "scenario", "scenarios")