

UNIVERSITY OF LJUBLJANA  
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

**THE CONNECTION BETWEEN SUSTAINABILITY REPORTING  
AND FINANCIAL PERFORMANCE IN SLOVENIAN COMPANIES**

Ljubljana, January 2023

JAN KOLENC

## AUTHORSHIP STATEMENT

The undersigned Jan Kolenc, a student at the University of Ljubljana, Faculty of Economics, (hereafter: SEB LU), author of this written final work of studies with the title The connection between sustainability reporting and financial performance in Slovenian companies prepared under the supervision of Vesna Žabkar Ph.D

### DECLARE

1. this written final work of studies is to be based on the results of my research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies is to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and/or quoted works and opinions of other authors in this written final work of studies per the FELU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted per the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on this written final work could have for my status at the FELU per the relevant FELU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have marked them;
7. to have acted per the ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained the permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to the publication of my personal data is included in this written final work of studies and this declaration when this written final work of studies is published.

Ljubljana, January 4<sup>th</sup>, 2023

Author's signature: \_\_\_\_\_

# TABLE OF CONTENTS

<b>INTRODUCTION</b> .....	<b>1</b>
<b>1 THE CONCEPT OF SUSTAINABILITY reporting</b> .....	<b>3</b>
<b>1.1 Sustainability reporting</b> .....	<b>3</b>
<b>1.2 Sustainability ratings</b> .....	<b>6</b>
<b>1.3 Sustainability reporting in Slovenia</b> .....	<b>10</b>
<b>1.4 Connection between sustainability and financial performance</b> .....	<b>11</b>
<b>2 RESEARCH FRAMEWORK AND METHODOLOGY</b> .....	<b>13</b>
<b>2.1 Research Goals and Research Questions</b> .....	<b>14</b>
<b>2.2 Methodology and Research Process</b> .....	<b>15</b>
2.2.1 Sustainability report analysis .....	16
2.2.1.1 <i>Data sets and OCR</i> .....	17
2.2.1.2 <i>Cleaning the data</i> .....	19
2.2.1.3 <i>Lemmatization</i> .....	20
2.2.1.4 <i>AFINN</i> .....	21
2.2.1.5 <i>Readability</i> .....	22
2.2.1.6 <i>TF-IDF</i> .....	22
2.2.1.7 <i>LDA</i> .....	23
2.2.2 Financial performance indicators .....	24
2.2.2.1 <i>Selection of indicators and calculations</i> .....	24
2.2.2.2 <i>Descriptive statistics of indicators</i> .....	26
<b>3 ANALYSIS AND RESULTS</b> .....	<b>26</b>
<b>3.1 AFINN</b> .....	<b>29</b>
<b>3.2 Readability</b> .....	<b>30</b>
<b>3.3 Correlation analysis</b> .....	<b>31</b>
<b>3.4 LDA</b> .....	<b>32</b>
<b>3.5 Regression models</b> .....	<b>35</b>
<b>3.6 Regression assumptions</b> .....	<b>36</b>
3.6.1 Independence of observations .....	36

3.6.2	No Hidden or Missing Variables .....	37
3.6.3	Linear relationship .....	37
3.6.4	Normality of the residuals.....	37
3.6.5	No or little multicollinearity .....	38
3.6.6	Homoscedasticity .....	38
3.6.7	All independent variables are uncorrelated with the error term.....	39
3.6.8	No correlation between the error terms .....	39
<b>3.7</b>	<b>Regression results .....</b>	<b>40</b>
<b>4</b>	<b>DISCUSSION.....</b>	<b>47</b>
<b>4.1</b>	<b>Interpretation of main findings and conclusions.....</b>	<b>47</b>
<b>4.2</b>	<b>Limitations and future research .....</b>	<b>49</b>
	<b>REFERENCE LIST.....</b>	<b>50</b>
	<b>APPENDICES .....</b>	<b>55</b>

## LIST OF FIGURES

Figure 1:	The 17 Sustainable development goals.....	4
Figure 2:	Timeline of CSRD implementation .....	5
Figure 3:	ESG scores for some of the biggest companies in the world.....	8
Figure 4:	Sub-categories of the ESG pillars by S&P Global.....	9
Figure 5:	Preparedness fields by S&P Global .....	9
Figure 6:	Scanned report.....	17
Figure 7:	Report filled with info-graphics .....	18
Figure 8:	Input (left) and output (right) of tesseract.....	18
Figure 9:	R code to perform OCR .....	19
Figure 10:	R code to perform the corpus cleaning .....	20
Figure 11:	Lemmatization example in Python .....	21
Figure 12:	LDA process .....	24
Figure 13:	Number of sustainability reports by industries .....	27
Figure 14:	Average length of sustainability reports by industries.....	28
Figure 15:	Average length of sustainability reports through years (words).....	28
Figure 16:	Average ratio of sustainably marked words to others by industries .....	29
Figure 17:	Average sentiment of sustainability reports through the years .....	30
Figure 18:	Average readability of sustainability reports through the years .....	30
Figure 19:	Clusters of words representing topics .....	31
Figure 20:	LDA analysis predicting 3 themes on the whole dataset .....	32
Figure 21:	LDA analysis predicting 3 themes in the year 2014 for all companies .....	33

Figure 22: LDA analysis predicting 3 themes in the year 2020 for all companies.....	34
Figure 23: LDA analysis predicting 3 themes in industry 3 for all years.....	35
Figure 24: LDA analysis predicting 3 themes in industry 9 for all years.....	35
Figure 25: Normal Q-Q plot .....	37
Figure 26: Homoscedasticity plot.....	39
Figure 27: Residual autocorrelation plot .....	39

## LIST OF TABLES

Table 1: Literature overview.....	13
Table 2: Methods used in the analysis .....	16
Table 3: Example of the lemmatization with lemmagen .....	20
Table 4: Example of the updated AFINN table .....	21
Table 5: Example of the sentiment calculation.....	21
Table 6: Reading difficulty .....	22
Table 7: Financial performance indicators .....	24
Table 8: Financial indicators descriptive statistics .....	26
Table 9: Industry classification.....	26
Table 10: VIF for each of the independent variables.....	38
Table 11: Regression results for financial indicators of PROFITABILITY .....	40
Table 12: Regression results for financial indicators of MARGINS.....	41
Table 13: Regression results for financial indicators of INDEBTEDNESS .....	44
Table 14: Regression results for financial indicators of LIQUIDITY AND SOLVENCY .....	45
Table 15: Regression results for regressing company sector on quality variables .....	46
Table 16: Results Summary .....	49

## LIST OF APPENDICES

Appendix 1: Povzetek (Summary in Slovene language).....	1
---	---

## LIST OF ABBREVIATIONS

CSRD – Corporate Sustainability Reporting directive  
 GRI – global reporting initiative  
 SFDR – Sustainable Finance Disclosure Regulation



## INTRODUCTION

In recent years there has been considerable traction in comparing financial performance and sustainability reporting of companies worldwide. The lack of research in this field in Slovenia is one of the main drivers for writing this thesis. Before heading into details, it is fair to briefly present the topic and the factors that have been the motivation for choosing this topic.

The word sustainable originates from the Latin *sustinere*, meaning "to stand firm" (WordSense Dictionary, n.d.). In business, it refers to ensuring long-term success, while maintaining or enhancing the state of the environment. Sustainability is about ensuring that there will be purity of water, air, and land and that there will be no scarcity of food, thereby to meet our basic human needs now and in any time in the future.

Nowadays, to make money, market participants are pressured by the public and different interest groups, to think about how to succeed in a sustainable manner. Besides, leaning toward sustainability provides self-fulfilment to managers and company employees and delivers positive externalities for the society. These positive effects are taken for granted to such an extent that they are not only being encouraged, but the market is punishing the lack of them. Investors are prepared to step away or even encourage divesting in a company if they deem it unsustainable.

Divestment and similar techniques are being used more and more often to push companies of all sizes toward a sustainable business model, with a sense of environmental and social impact (Pizzi, Corbo & Caputo, 2021). A notable party, exerting pressure towards achieving sustainable businesses, are activist hedge funds.

One of the greatest "battles" for a better tomorrow in the sense of environment was between Exxon Mobil<sup>1</sup> and Engine No. 1<sup>2</sup>. The latter being an activist hedge fund that entered into a proxy contest<sup>3</sup> with the board of Exxon Mobil. The contest was started due to claims that the company was lagging in cleaner fuels and its financial underperformance, compared to its competitors, leading to over 22 billion USD in losses. The activists spent over 65 million USD to get seats on the board of directors at Exxon, in order to direct the company towards a greener path. Such efforts not only encourage sustainable behaviour in other companies, but also strike the fear of hostile takeovers in some passive managers (Hiller & Herbst-Bayliss, 2021).

In order to avoid action, companies similar to the ones mentioned above need to convince investors and the broader society that attempts toward a greener tomorrow are being made,

---

<sup>1</sup> Exxon Mobil one of the largest oil and gas companies.

<sup>2</sup> Engine No. 1 is an American activist hedge fund

<sup>3</sup> »A proxy contest is a tactic frequently used by a party attempting a hostile takeover or a dissident group of stockholders to replace a company's existing board of directors with new directors that support the potential acquirer's or dissident stockholders' agenda and objectives.« (Thomson Reuters, n.d.).

by providing efficient and accurate sustainability reports. These reports provide a great source of insight and are, to some extent standardized, to provide an option for inter-company comparisons.

Besides the reports, insight into the efforts on sustainability issues of companies, is also provided by sustainability ratings. Those are defined as: "numerical scores, percentages, or letter grades, that aim to provide a snapshot of an entity's exposure to environmental, social, and governance risks, and how effectively it manages those risks." (Rio ESG, 2021) Using scores allows for inter-company benchmarking and additionally, more straightforward analysis for links between the financial and sustainability aspect of a company.

Following the above, a question arises whether sustainability reports offer a valuable insight into the financial performance of companies and moreover, how to efficiently and fairly measure it between companies. The main focus of this master thesis is therefore to discover whether technical analysis of sustainability reports quality provides a useful tool in finding the link between financial performance and sustainability reports.

The link was sought after by the answers to the following research questions:

1. To what extent do a company efficient and well-structured sustainability reports relate to good financial performance?
2. What can one deduce from the contents of a sustainability report and its properties about future financial performance?
3. How does sustainability report quality depend on the firm sector?

The master thesis is structured in the following way to answer these research questions.

The first chapter of the master thesis describes the idea of sustainability reporting. Discussion is conducted in a way that provides definitions and introduces the essential companies and organizations participating in the field of sustainability reporting. The following section focuses on sustainability scores and introduces many prominent companies that provide sustainability scores. One of the companies is highlighted to clarify the general approaches for calculating a company sustainability score. In order to conclude the oversight section of the first chapter in a meaningful way, the subchapter 1.3 concentrates on presenting the existing and potential future legislation governing sustainability reporting in Slovenia. The concluding subchapter of chapter 1 provides the findings of several academic papers on the topic.

In the following part, the emphasis is placed on presenting the research questions and describing the strategies utilized to address them. The first subchapter gives more insight into the subject matter of papers in the area and illuminates the void in the literature that this article addresses. The succeeding section provides a more thorough explanation of the methods used and clarifies some operations with examples.



The technical analysis is conducted primarily in R and somewhat in Python. The analysed documents comprise of around 702 reports, dated between 2014 and 2020, from approximately 130 firms. OCR is the first method used on the company sustainability reports in order to convert them to text. To identify sentiment, a technique and library created by Finn Arup Nielsen (AFINN), that has been translated to Slovene by Martinc and then adapted to analyse sustainability reports, were used. Quanteda,<sup>4</sup> textstats, and the Flesch–Kincaid<sup>5</sup> readability tests were utilized to measure readability. The TF-IDF<sup>6</sup> approach was used to identify the report-specific terms, to avoid utilizing, since they may influence the findings. LDA, an abbreviation for latent Dirichlet allocation, produced subjects developed in business reports or groupings of company reports.

Regression analysis was used on the above metrics and financial performance indicators, to answer the research questions and establish relationships. Financial information was acquired from AJPES<sup>7</sup> in order to determine each company's financial performance metrics.

Chapter 3 describes the results of the research methods from the previous chapter. The results are presented in tables and figures that depict the findings. The results are then discussed in the last chapter, where the focus is on answering the research questions and bringing forward the caveats and possibilities for future research.

## **1 THE CONCEPT OF SUSTAINABILITY REPORTING**

### **1.1 Sustainability reporting**

Sustainability reporting is a tool for showing the market how aware, concerned and active a company is, regarding three pillars of ESG: environmental, social and governance. Nowadays, approximately 30% of all the assets on the world markets are subject to ESG standards. Since 2016, the number has increased by more than a third. These figures, represent a rising understanding among businesses, financiers, and stockholders, that in order to be successful, organizations should give increasingly more thought and effort to their environmental footprint in novel ways (Howard-Grenville, 2021).

The leading institution, pushing companies towards standards, set with ESG, are the United Nations, with their enactment of the United Nations' Sustainable Development Goals (SDGs) in 2015. They call for all developed and developing nations to join in a global partnership to pursue common goals. They indicate that eradication of poverty and other forms of deprivation should be pursued in conjunction with initiatives that enhance

---

<sup>4</sup> »A fast, flexible, and comprehensive framework for quantitative text analysis in R« (Benoit et al., 2022)

<sup>5</sup> Readability test to determine how hard the text is to read on a certain scale.

<sup>6</sup> »term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus.« ('Tf–Idf', 2022)

<sup>7</sup> Agencija Republike Slovenije za javnopravne evidence in storitve (Agency of the Republic of Slovenia for Public Legal Records and Related Services).

wellbeing and education, reduce inequality, fuel economic development, and address sustainability (United Nations, n.d.). The 17 SDGs are depicted in Figure 1. Apart from the close relation between SDGs and ESG, there is often confusion, when discussing ESG and the triple bottom line approach (TBL). Despite having matching objectives, ESG and triple bottom line are two distinct concepts. The triple bottom line is a paradigm for accounting that tries to widen the prominence of corporations beyond P&L,<sup>8</sup> namely to incorporate social and environmental concerns. TBL consists of three Ps: profit, people, and the planet. The primary distinction between the TBL and ESG is that the first emphasizes business social and environmental fundamentals in addition to profit. ESG is more fixated on impacting investment decisions by taking environmental, social and governance principles into account (ESG – The Report, n.d.).

Figure 1: The 17 Sustainable development goals



Source: United Nations (n.d.).

In Europe, the most significant turns toward reporting on sustainability were done by the European Parliament, by adopting the proposal for the Corporate Sustainability Reporting directive (hereafter: CSRD) (Circularise, 2022).

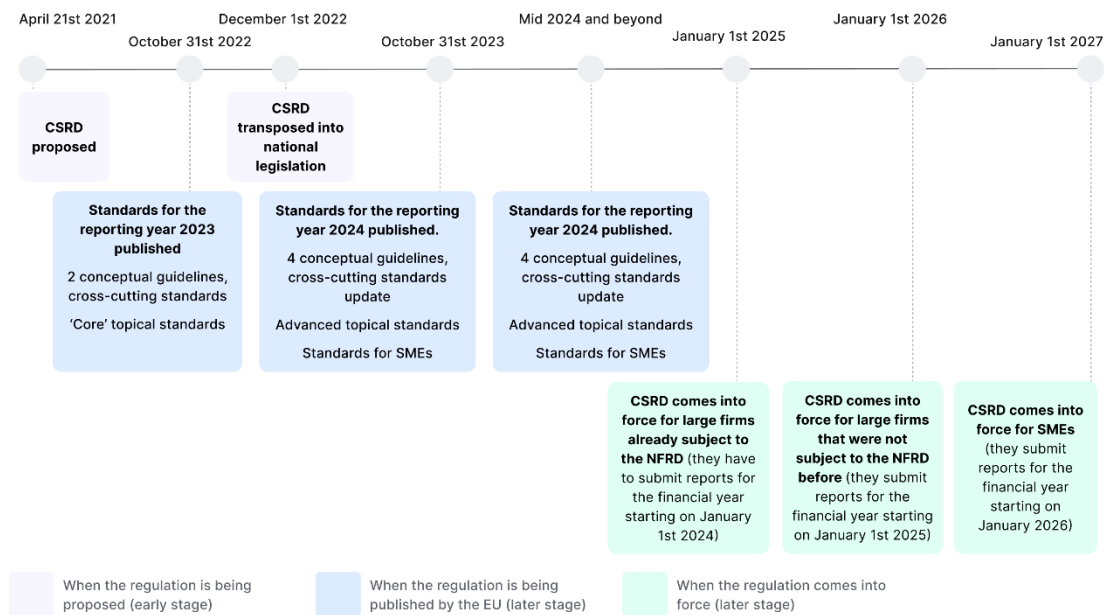
By establishing a standardized reporting structure, the CSRD elevates the level and calibre of sustainability data. The legislation's ultimate goal is to make the enterprises more credible in the eyes of all parties, including banks, investors, clients, and consumers (Circularise, 2022).

---

<sup>8</sup> Profit and loss.

On April 21st, 2021, the text of the proposal was availed. Member states must incorporate the directive into their national laws by December 1st 2022. For big enterprises, which are already subject to the NFRD<sup>9</sup>, the directive takes effect in 2025; for large firms not yet covered by the rule, it does so a year later. The CSRD also applies to SMEs<sup>10</sup> as of 2027. The chronology is shown in more detail in Figure 2 below (Circularise, 2022).

Figure 2: Timeline of CSRD implementation



Source: Circularise (2022).

The European Union intends to align the standards with current sustainability frameworks, international efforts for sustainability reporting, and the EU's regulatory framework (the European Green Deal (hereafter: EGD), Sustainable Finance Disclosure Regulation (hereafter: SFDR), and Taxonomy Regulation).

The EGD, unveiled in December 2019, proposes an outline for initiatives to increase resource efficiency by transitioning to a clean and circular economy, halting climate change, contrain loss of biodiversity, and diminish pollution. It describes the compulsory investments and obtainable finance options and discusses, how to assure an equitable and all-encompassing alteration. According to estimates, achieving the objectives for 2030, would need an annual investment increase of 260 billion Euros. This investment will need governmental and private sector deployment in sense of assets and action (European Commission, 2019).

The significance and the ideas behind the EGD are well portrayed with the following statement of Ursula von der Leyen: “The European Green Deal is our new growth strategy

<sup>9</sup> The Non-Financial Reporting Directive.

<sup>10</sup> Small and medium enterprises.

– for a growth that gives back more than it takes away. It shows how to transform our way of living and working, producing and consuming so that we live healthier and make our businesses innovative.” (European Commission, 2019)

The EU SFDR is an assortment of EU regulations devised to help investors more straightforwardly understand and compare the sustainability features of funds. As its name suggests, a greater emphasis is placed on transparency, including implementing fresh regulations, mandating disclosure of any adverse consequences created by investee enterprises. The SFDR primarily applies to EU-based financial institutions. Indirectly, non-EU companies are affected by their subsidiaries in the EU, the provision of services in the EU, and market forces. The SFDR requires financial market participants and advisors to disclose information about the sustainability aspects of both ESG and non-ESG products. The rule mandates that businesses identify the items or the guidance they provide into three categories: mainstream products, products promoting environmental or social characteristics, and products with sustainable investment objectives (PricewaterhouseCoopers, n.d.).

The EU taxonomy is a scheme that classifies environmentally sustainable commercial activities. The EU taxonomy could be a key factor in helping the EU increase sustainable investment and achieve its goals under the European Green Deal. It would provide businesses, investors, and regulators with clear definitions for identifying environmentally sustainable economic activities. In this manner, it should provide investors with security, shield private investors from greenwashing and assist businesses in becoming more sustainable (European Commission, n.d.).

As a result of aligning CSRD with the above-explained building blocks of the EU regulatory framework, comparable metrics will be included in EU sustainability reports. The standards will cover the information that SFDR-compliant financial market participants are required to disclose. The companies will assess the technical screening standards, outlined in the Taxonomy Regulation and its delegated acts to determine, if their operations are viable. The influence of sustainability concerns on company performance, position, and growth, as well as the effects of its operations on people and the environment must be included in the report (Circularise, 2022).

As noted before, sustainability reporting is an excellent indicator for the market about a company's commitment for a sustainable business strategy. However, this is not the sole source of information of the latter. Sustainability ratings developed by analytic businesses are an excellent measure of company sustainability.

## **1.2 Sustainability ratings**

Creating ESG data analytics, for assessing ESG activities and effects, is the key to realizing sustainability. The idea is that businesses will feel pressured into performing better, by

making such metrics public, via a reliable scoring system. These evaluation techniques would present procedures, applicable to all companies, that would result in a score that could be used to compare the efforts of individual companies. It is challenging to create an algorithm that would determine a numerical score for, by its nature, a very descriptive dataset (Howard-Grenville, 2021).

ESG data analytics is the procedure of gathering, storing, and evaluating the information on the ESG performance of an organization. This information may be obtained through several sources, including financial records, public releases, and media coverage. ESG data analytics can be used to calculate algebraic values, known as ESG metrics, that assess the environmental, social, and governance performance of a company. Businesses can use machine learning algorithms to detect ESG risk factors and predict their potential future effect. This information may then be utilized, to make strategic resource distribution and risk moderation adoptions (Kumar, 2022).

Apart from the challenge presented by the evaluation of sustainability activities, there is also a barrier, when defining the issues about sustainability, namely, to what extent a company is responsible for its actions, and actions caused by third parties, using products of that company.

For example, company X is producing a product, for which the company produces Y of negative externalities. A third party then uses this product, and while using the product, the third party produces Z negative externalities. Should Company X, in terms of ESG, be held responsible for Y or Y+Z negative externalities?

To solve such dilemmas, standardising ESG standards and scoring procedures are of utmost importance to provide fair treatment to all parties.

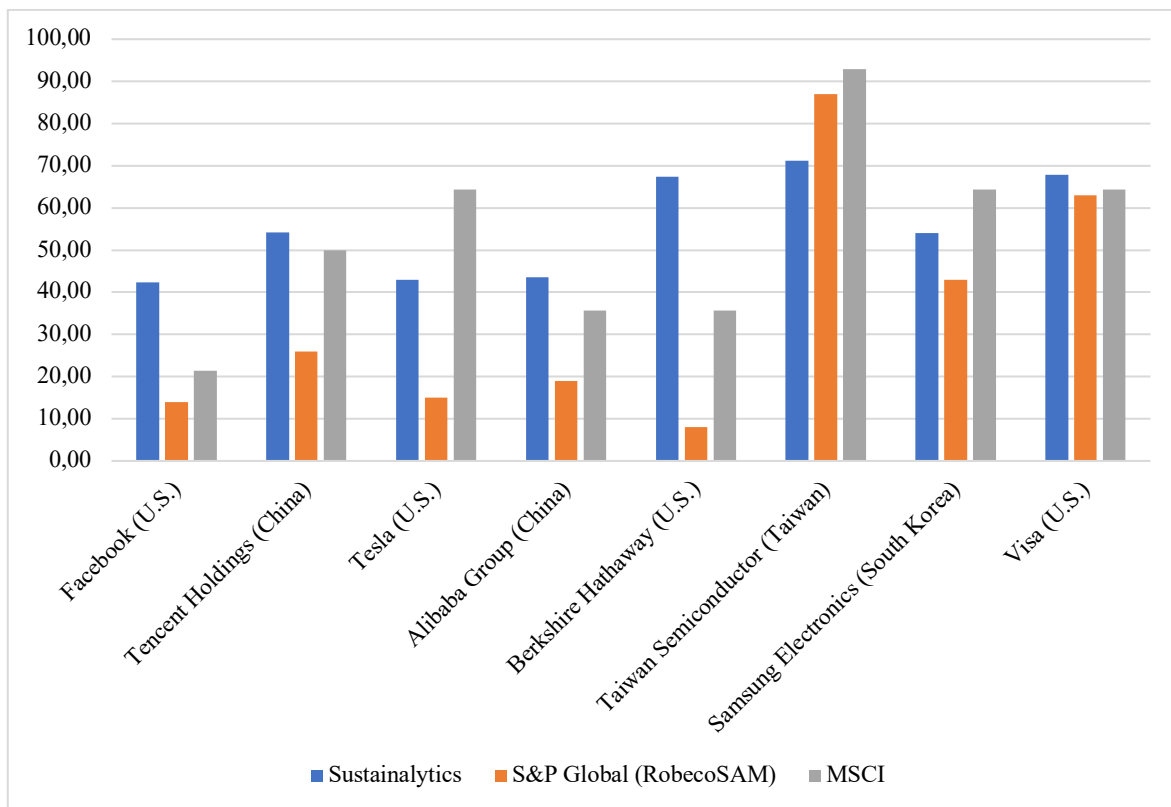
The global reporting initiative (hereafter: GRI) is an organization, that has been active in the field of sustainability reporting form 1997. Their reporting standards are designed to be applied wherever and to whichever company. "The GRI Standards enable any organization – large or small, private or public – to understand and report on their impacts on the economy, environment and people in a comparable and credible way, thereby increasing transparency on their contribution to sustainable development. In addition to reporting companies, the Standards are highly relevant to many stakeholders – including investors, policymakers, capital markets, and civil society" (Global Reporting Initiative, n.d.).

The GRI standards are widely used in research, when trying to asses the sustainability of a company through their nonfinancial report. Researchers, with the help of GRI standrards, create scores, which they use in analysis, when comparing sustainability and financial performance of a company.

Measuring of ESG company performance is also done by analytical companies, who have internally set out rules for measurement, but are not yet synced amongst the competitors. In

Figure 3, we can examine the ESG ratings of some of the world significant firms as determined by the three leading ESG score suppliers. As shown in the case of the Alibaba group, the S&P Global score is greater than the S&P Global score for Berkshire Hathaway, but the Sustainalytics score for Berkshire Hathaway is higher than the score for the Alibaba group. Interestingly, the MSCI scores for both firms are identical.

*Figure 3: ESG scores for some of the biggest companies in the world*

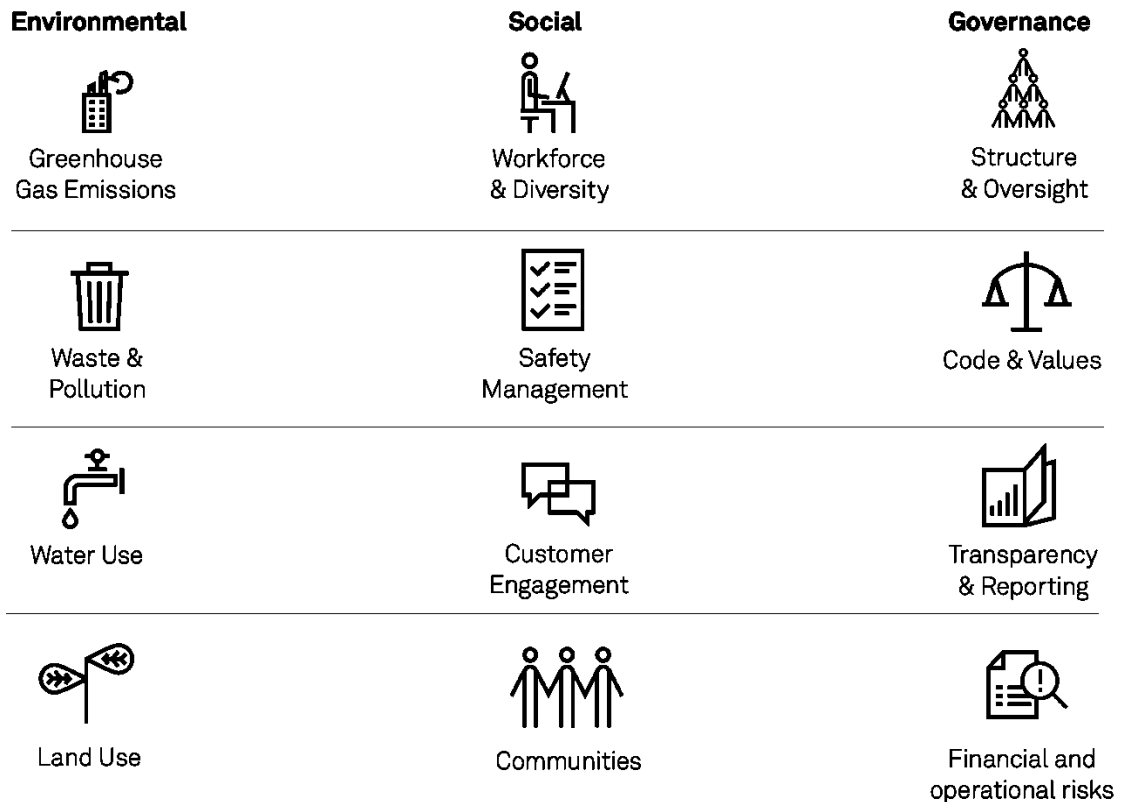


*Source: Statista (2022).*

S&P Global is one of the companies competing in sustainability ratings, as seen before. They are one of the six significant participants in the industry (the others being Institutional Shareholder Services (ISS), Morgan Stanley Capital International (MSCI), Sustainalytics, and FTSE Russell). To illustrate the method of producing an ESG score more clearly, since the approach may be generalized, I have picked just one of the companies.

The ESG evaluation provided by S&P Global consists of two primary partitions. The first is measuring the entity's capacity to manage its risk exposure and identify opportunities compared to peers over the short to medium term. The score incorporates the evaluation of three pillars of ESG. The subdivision of the pillars can be seen in Figure 4. These categories, when combined, provide a benchmark for comparison across sectors and regions. The assessment is done with the use of a questionnaire, and when necessary, completed by additional questions from analysts for further clarification, in order to provide a fair scoring (Rio ESG, 2021).

Figure 4: Sub-categories of the ESG pillars by S&P Global



Source: S&P Global (n.d.).

The second part of the ESG evaluation comes from preparedness<sup>11</sup>. Analysts consult with top management and firm's board members to obtain a clear view about their perceptions of new trends, possible competitors, and long-term strategy. The main fields being covered by measuring preparedness are portrayed in Figure 5 below.

Figure 5: Preparedness fields by S&P Global



Source: S&P Global (n.d.).

<sup>11</sup> Preparedness or preparedness opinion is a qualitative assessment of a company's ability to foresee and respond to adverse events in the long term (S&P Global, n.d.).

The following quotation from Pinebridge Investments' Alessia Falsarone, Head of Sustainable Investing, highlights the significance of including preparation in the ESG assessment: "The Preparedness Opinion is particularly relevant to my risk lens. It builds on understanding the nexus between the organizational awareness journey of a company, its impact on culture and how it translates into what I call coherent, repeatable and verifiable processes." (S&P Global, n.d.)

Even if the ESG ratings could be helpful, the focus of this study is on sustainability reports. I will closely inspect the sustainability reports of some of the most prominent Slovenian businesses. I will create measurements and evaluate their generalizability compared to ESG grading techniques. The metrics will consist of readability, sentiment analysis, length of reports and derivatives of these metrics combined; they comprise quantifiers for the quality of sustainability reports. All of the metrics and procedures of calculation are explained in chapter 2.2. I will concentrate on Slovenian sustainability reports, which became more widely available after 2013 and now comprise a sufficiently large database for analysis.

### **1.3 Sustainability reporting in Slovenia**

In Slovenia, the legal framework for reporting on nonfinancial aspects (sustainability) of companies' operations was provisionally prescribed by the EU in Directive 2013/34 and Directive 2014/95. The directives were incorporated into Slovenian law by an amendment to the Companies Act, specifically in the provisions on the business report, as part of the annual reporting of companies (Trgovinska zbornica Slovenije, 2022; Zakon o gospodarskih družbah, ZGD, Uradni list RS, št. 65/09 – uradno prečiščeno besedilo, 33/11, 91/11, 32/12, 57/12, 44/13 – odl. US, 82/13, 55/15, 15/17, 22/19 – ZPosS, 158/20 – ZIntPK-C in 18/21).

Currently, only firms of public interest with an average of more than 500 employees, are obliged to submit a nonfinancial activity statement. Companies having a consolidated average of more than 500 employees are also obliged to submit a consolidated annual report, including a part on sustainability. Therefore, the law stipulates that these companies' annual reports accurately depict the company, describe significant risks, and provide other information pertinent to comprehend the business and its progress. In doing so, the law emphasizes that this information also refers to protection of environment and safety of personnel (Trgovinska zbornica Slovenije, 2022; ZGD).

In general, Slovenian legislators adhere to European Union suggestions. The suggestion in section 1.1 broadens the scope to cover all market-listed corporations. It mandates examining the provided information, more extensive reporting requirements, and compliance with statutory EU sustainability reporting criteria. In addition, the idea compels businesses to digitally label their provided data as machine-readable (Trgovinska zbornica Slovenije, 2022).



The proposal for a regulation on sustainable corporate reporting mandates reporting standards at the European level. It is anticipated that the European Financial Reporting Advisory Group will develop the proposed standards (EFRAG). (Trgovinska zbornica Slovenije, 2022). In addition to conforming to EU regulations, the standards will continue to support worldwide standardization projects. It is anticipated that by October 2022 the first criteria for sustainability reporting will be established (Trgovinska zbornica Slovenije, 2022).

#### **1.4 Connection between sustainability and financial performance**

The research included below, examined the relationship between sustainability reporting and financial success in various approaches. Some have discovered favourable relationships, some negative, and others were inconclusive. I will corroborate on some articles that I felt were most pertinent and applicable to my situation, and in Table 1, at the end of the chapter, I will sum up the findings and metrics of the discussed articles.

Ching, Gerab and Toste (2017) have conducted research using a database of sustainability reports from 2008 to 2014 from Brazilian enterprises. During this time frame they have amassed 218 reports from 51 distinct businesses in Brazil. Using the GRI frameworks, they have studied the quality of reports. The final score was determined by including a 0 to 1 score for the quantity of information disclosed in the report. In order to evaluate financial performance, the following metrics were analysed: Capex, financial leverage, ROA, ROE, net margin, price to book value, market capitalization, earnings per share, total assets, and operational cash flow. These factors were then assessed by means of linear regression to define their relationship to sustainability. They determined that there was no association amid accounting and market-based characteristics and reporting quality and that despite the fact that the quality of disclosure has enhanced throughout the years observed the scores persist low. They were unaware of any studies analysing the association between financial success and the quality of sustainability reporting; consequently, their study was the only one of its kind.

Loh, Thomas and Wang (2017) have prepared their research on 502 reports of companies that were listed on the SGX Mainboard (Singapore Exchange). Using a measuring approach, created by the ASEAN CSR Network and Centre for Governance, Institutions, and Organizations, they deliver a score to assess the degree of sustainability reporting. There were 23 criteria in the program, organized into four indicators: Governance, Economic, Environmental and Social. A score between 1 and 5 was specified for each criterion and then translated to the indicator's score with equal weighting, such that the maximum score for each indicator (Governance, Economic, Environmental and Social) is 25. Ultimately, the aggregate of the four indications provided a total score ranging from 20 to 100 points. For the financial indicators they chose: market value, four months after the financial end of the year, the book value of common equity at the company's end of the year and earnings before

extraordinary items at the company's end. After conducting linear regressions, they concluded that sustainability reporting is significantly linked to a firm's market value and this connection is impartial to sector or company status, such as government-affiliated or family-owned companies.

Laskar (2019) explored the link between sustainability reporting and the profitability of Indian and South Korean corporations. The sample involved 28 nonfinancial establishments from India and 26 nonfinancial businesses from South Korea between 2010 and 2015. A qualitative approach based on the GRI was applied to determine the sustainability performance score. Regression was used to analyse the relationship between sustainability reporting and company performance using the disclosure ratings. The regression findings revealed that the correlation is positive and statistically significant for South Korean businesses. However, in the Indian context, sustainability performance has a negative influence; the authors explain that this might be due to the fact, that sustainability is a relatively new concept in Indian companies and that it presents financial costs or reallocation of resources for the companies. In addition, the proportional effect of sustainability reporting is substantially more significant in South Korea than in India. Similarly (Buallay, Hamdan & Barone, 2019) could not come to a conclusive answer when searching for a link between financial performance and sustainability.

In their work, Girón, Kazemikhasragh, Cicchiello and Panetti (2021) explored the link between reporting activity and economic success of enterprises, but their primary objective was to examine the variables that impact the employment of new sustainability reporting standards and external assertion. The paper incorporates evidence from two databases, notably the GRI Sustainability Disclosure Database and the Orbis database, on 366 significant Asian and African businesses that have addressed the SDGs in their 2017 sustainability reports. Their results indicate that taxes, assets and profit margin substantially affect sustainable development goals and economic performance as evaluated by Tobin's q. The findings also indicate that the manufacturing industry positively correlates with the company's choice to report using SDG frameworks. There is a significant association between capital and leverage and their dependent variables.

Slovenian authors (Ermenc, Klemenčič & Rejc Buhovac, 2017) involved 79 nonfinancial Slovenian enterprises in their sample. Between 2007 and 2014, financial information was collected for the companies in scope of the research. In 2011, data regarding sustainability, was available for 44 companies in their annual report. The remaining corporation's corporate sustainability was assessed, based on their 2010 annual reports. For approximating corporate sustainability, they have examined company annual reports using a content analysis approach, developed by (Slapničar, 2004). Environmental management, interaction with the local community, and relationship with workers, suppliers, purchasers, and availability of an annual report, comprised the index measuring corporate sustainability. To measure financial performance, they have focused on ROA. The study conducted regressions over different time periods and found a positive relationship between higher levels of corporate

sustainability and financial success. However, they did not find a statistically significant correlation concerning a company's average historical financial success and its succeeding sustainability performance.

Hussain (2015); Kasbun, Teh and Ong (2016); Motwani and Pandya (2016); Al-Wattar, Almagtome and Al-Shafeay (2019) and Nguyen (2020) all provided verification of a positive correlation between sustainability reporting and financial performance. They all had the focus on ROA and some case-specific indicators in common (Wahyuni, 2020). Asuquo, Temitayo and Raphael (2018) could not find comparable results as the group above, as they could not prove the correlation between financial performance and sustainability reporting.

For easier oversight, the results of all of the above-mentioned studies are summarized in Table 1 below.

*Table 1: Literature overview*

<b>Article</b>	<b>ESG variables</b>	<b>Financial performance variables</b>	<b>Years</b>	<b>Country</b>	<b>Result</b>
(Ching, Gerab & Toste, 2017)	GRI approach + own 0-1 scale	Capex, financial leverage, ROA, ROE, net margin %, price to book value, market capitalization, earnings per share, total assets, and operational cash flow	2008-2014	Brazil	+
(Loh Thomas & Wang, 2017)	ASEAN CSR on a scale of 20-100	Market value, book value, earnings	-2016	Singapore	+
(Laskar, 2019)	GRI approach	Profitability	2010-2015	South Korea /India	+/-
(Girón, Kazemikhasragh, Cicchiello & Panetti 2021)	GRI approach	Taxes, assets, profit margin, Tobin's q	2017	Asia/Africa	+
(Ermenc, Klemenčič & Rejc Buhovac, 2017)	Sustainability index by Slapničar	ROA	Fin: 2007-2014 ESG:2011	Slovenia	+
(Buallay, Hamdan & Barone, 2019)	Bloomberg indices	return on assets, return on equity and Tobin's q	2008-2017	Global	+/-
(Asuquo, Temitayo & Raphael, 2018)	GRI approach	ROA	2012-2016	Nigeria	-

*Table continues*

Table 1: Literature overview (cont.)

Article	ESG variables	Financial performance variables	Years	Country	Result
(Hussain, 2015)	GRI approach	firm size, capital intensity, size, Tobin Q, ROA, Debt to Equity ratio.	2007-2011	Global	+
(Kasbun, Teh & Ong, 2016)	GRI approach	ROE, ROA, Company size	2006-2013	Malaysia	+
(Motwani & Pandya, 2016)	GRI approach	ROA, ROE, ROCE, PBT and GTA	2009-2015	India	+
(Al-Wattar, Almagtome & Al-Shafeay, 2019)	GRI and Disclosure	ROA, ROS	2013-2018	Iraq	+
(Nguyen, 2020)	GRI approach	ROA, Firm size, Total debt / Total assets	2013-2017	Germany	+
(Wahyuni, 2020)	GRI approach	Price to Earnings Ratio (PER)	2016-2018	Indonesia	-

Source: Own work.

## 2 RESEARCH FRAMEWORK AND METHODOLOGY

### 2.1 Research Goals and Research Questions

In recent years there has been a great deal of interest in comparing global financial performance and sustainability reporting, which has led to the release of several works on this topic, mentioned below.

Researchers concentrate primarily on their own nation and the publicly traded enterprises operating in their local market. As seen in the above section (Literature review), these kinds of studies were conducted in Brazil, Iraq, the United States, Australia, New Zealand, Nigeria, Malaysia, India, and South Korea (Asuquo, Temitayo & Raphael, 2018; Ching, Gerab & Toste, 2017; Kasbun, Teh & Ong, 2016; Laskar, 2019; Al-Wattar, Almagtome & Al-Shafeay, 2019; Motwani & Pandya, 2016). In addition, Slovene writers (Ermenc, Klemenčič & Rejc Buhovac, 2017) have released a study titled "Sustainability Reporting in Slovenia: Does Sustainability Reporting Influence Financial Performance?" In contrast to my work, their research was limited to a single parameter developed by Slapničar, which consists of "environmental management, relationship with the local communities, and relationship with employees, suppliers, and buyers, as well as by the availability of the annual report." In contrast, I will analyse sustainability reports using a wide variety of parameters, combining

some previously employed techniques with some novel ones (Ermenc, Klemenčič & Rejc Buhovac, 2017).

Consequently, this master thesis aims to establish a possible connection between several nonfinancial sustainability reporting characteristics. I have employed several R programming methodologies and literature-referenced libraries to quantify them. In addition, using standard corporate performance indicators, I investigated the relationship between financial performance and sustainability reporting in companies that are required to provide sustainability reports.

The research questions are based on the above goals and are questioning the following:

1. To what extent do a company efficient and well-structured sustainability reports relate to good financial performance?
2. What can one deduce from the contents of a sustainability report and its properties about future financial performance?
3. How does sustainability report quality depend on the firm sector?

## **2.2 Methodology and Research Process**

The analysis of the master thesis will be conducted primarily in R and somewhat in Python. The corpus of analysed papers comprises around 702 reports, dated between 2014 and 2020, from approximately 130 firms that were prepared in the framework of the ARRS project P5-0128. The sustainability reports were obtained partially from AJPES and partially from websites of the companies that were considered in the research.

I first performed OCR on the company sustainability reports in order to convert them to text. The texts were subsequently organized into a corpus-type object. All of the texts were converted to lowercase and stripped of special characters. After that, they were lemmatized using a procedure created by IJS in language C and then transformed by Virag, Grabec, and Žejn into Python (GitHub, 2019–2020). To identify the sentiment, I used a technique and library, created by Finn Arup Nielsen (AFINN), that had been translated to Slovene by Martinc and then adapted for the purpose of analysing sustainability reports.

Quanteda, textstats, and the Flesch–Kincaid readability tests were used to measure readability; all will be explained in further detail in chapter 3.2.1.5. (2022's "Flesch–Kincaid Readability Tests"). Later, I used the TF-IDF approach, which helped me identify the report-specific terms, to avoid utilizing them, since they may influence the findings and analysis. This is further explained in chapter 3.2.1.5 (Wikipedia, 2022a). LDA, an abbreviation for latent Dirichlet allocation, supplied me with particular subjects developed in business reports or groupings of company reports and is further explained in chapter 3.2.1.7.

In addition, I have supplied basic report information, such as average sentence length, number of sentences, number of characters, and number of words. I have conducted regression analysis on these metrics and financial performance indicators in order to answer the research questions and establish relationships. Financial information was acquired from AJPES in order to determine each company's financial performance metrics; I have examined this process annually. All of the aforementioned approaches are described in more depth and illustrated with examples in the following sections.

### 2.2.1 Sustainability report analysis

In this section (and its sub-sections), I will focus on explaining the methodologies used to analyse sustainability reports. In the first sub-section the main focus will be on the data sets and OCR (optical character recognition). Then, in the second sub-section, the data cleaning process will be described. The lemmatisation process and its creators will be discussed in the third sub-chapter. Afterwards, I will explain the method for evaluating sentiment AFINN. In the last three sub-chapters, I will discuss and shed some light on readability, the TF-IDF method and lastly, latent Dirichlet analysis. All of the methods are summarized in Table 2.

*Table 2: Methods used in the analysis*

<b>Method</b>	<b>Process</b>	<b>R packages used</b>	<b>Chapter used</b>
OCR	The method used to transform machine-unreadable text to machine-readable.	Tesseract, pdftools	3.2.1.1
Data cleaning	A chain of transformations is used to standardize the format of the text. Paragraphs are joined into one. Text is transformed into lower case, stripped of punctuations except for dots and Slovene alphabet letters are changed into English	Stringr, NLP, tidyverse, tidytext	3.2.1.2
Lemmatization	A process of transforming words in their original form to the stem for more accessible further analysis.	(python package) lemmagen	3.2.1.3
AFINN	A process to determine the sentiment of a text. Words have pre-defined sentiment values that are then averaged to obtain the sentiment of the text as a whole.	(not an R package) AFINN table	3.2.1.4
Readability	A process to calculate an index that tells us how difficult a text is to read.	quantenda	3.2.1.5
TF-IDF	A method to determine which words are specific to only one of the documents to exclude them from further analysis later.	tidytext	3.2.1.6
LDA	A process to determine topics inside text documents.	Topic models	3.2.1.7

*Source: Own work.*

### 2.2.1.1 Data sets and OCR

The data set consisted of sustainability reports and annual reports of Slovenian companies, approximately 650, of the first contained in the latter. Altogether, the data set consists of 702 reports, from the year 2014 until the year 2020. Approximately 130 Slovenian companies delivered the reports.

The reports are of different qualities, as some are printed into complex form and then scanned to create a PDF file, which is handed to the authorities (as seen in Figure 6). On the other hand, some companies are putting extensive effort into their reports resulting in pages full of info-graphics and scattered text (as seen in Figure 7). For the sake of anonymization, I have redacted some texts that could have been used to identify the companies from the reports.

Figure 6: Scanned report

**Konti skupine 02 in 03 – Nepremičnine (AOP 004 in 005)**

v EUR, brez centov

konto	Naziv konta	
020	Zemljišča	
021	Zgradbe	
023	Nepremičnine v gradnji ali izdelavi	
02	Skupaj AOP 004	
03	Popravek vrednosti nepremičnin AOP 005	
02-03	Sedanja vrednost nepremičnin	

Nabavna vrednost nepremičnin – objektov se je v letu 2017 povečala za [redacted] EUR in je konec leta znašala [redacted] EUR. Povečanje je posledica zaključene investicije ureditev prostorov nove sterilizacije. Odpisana vrednost nepremičnin po obračunu amortizacije je [redacted] EUR, sedanja vrednost pa [redacted] EUR. V preteklih letih pričete in nezaključene investicije (helioport in 2. del energetske sanacije) predstavljajo [redacted] EUR.

61

---

V poslovnih knjigah so knjižene naslednje nepremičnine, ki na zadnji dan leta znašajo:

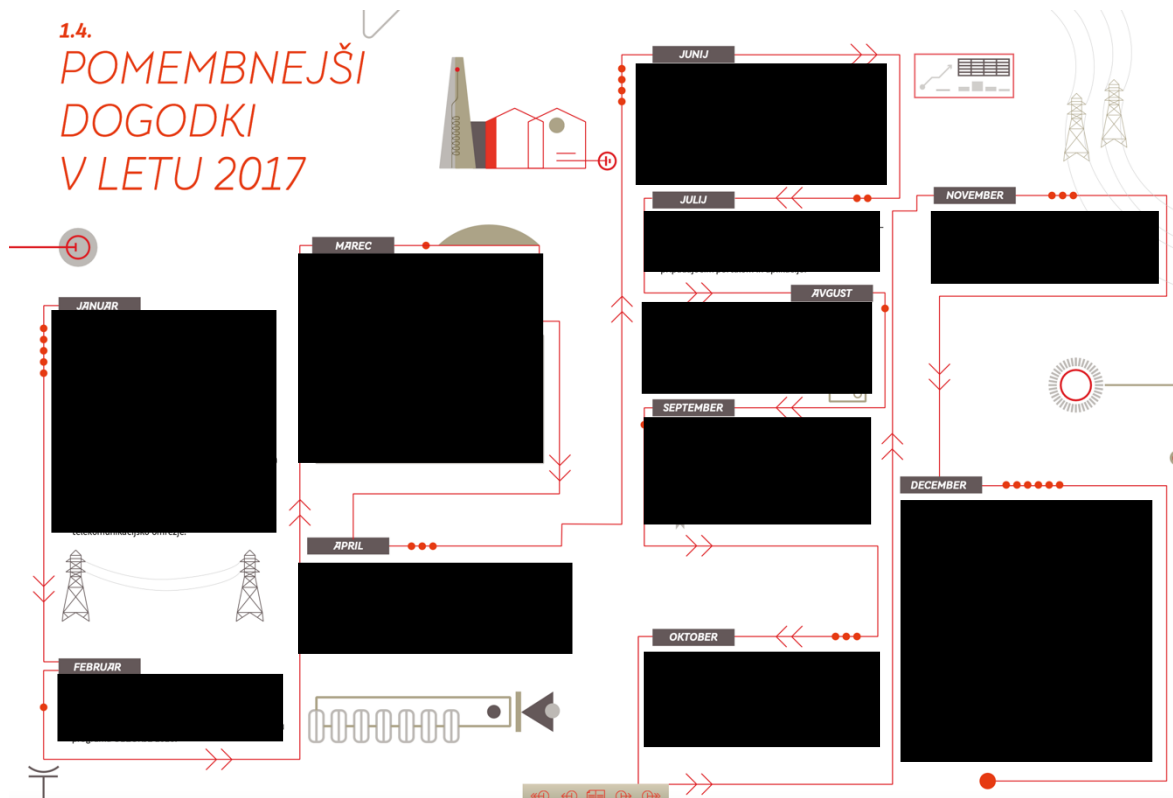
- zgradbe; [redacted] EUR,
- zemljišča ; [redacted] EUR.

Za nepremičnine imamo (delno še urejamo) zemljiško knjižno dokumentacijo, ki je kot kopija odložena v računovodski dokumentaciji in original v dokumentaciji ustanovitve pravne osebe-zavoda.

**Konti skupine 04 in 05 – Oprema in druga opredmetena osnovna sredstva (AOP 006 in 007)**

Source: Own work.

Figure 7: Report filled with info-graphics



Source: Own work.

The different quality of reports caused problems while importing the reports as text. In the first case, when the text was not machine-readable, the main issue was for R to recognize some special characters like letters ŠČŽ. There were also some issues with the text angles as it was not scanned wholly levelled. These problems were solved with the use of a package, called tesseract. This is a package that allows precise OCR on documents that are not machine-readable. This procedure is done on .png file types, so all the PDF documents had to be split into single pages, to be transformed into text by tesseract. Below in figure 8, an example of the input and output of tesseract can be seen.

Figure 8: Input (left) and output (right) of tesseract

This is a lot of 12 point text to test the ocr code and see if it works on all types of file format.  
 The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox.

```
## This is a lot of 12 point text to test the
## ocr code and see if it works on all types
## of file format.
##
## The quick brown dog jumped over the
## lazy fox. The quick brown dog jumped
## over the lazy fox. The quick brown dog
## jumped over the lazy fox. The quick
## brown dog jumped over the lazy fox.
```

Source: Institute for Statistics and Mathematics (n.d.).



The process of converting PDFs to PNG files is computationally intensive, so applying it to all of the reports would be inefficient. Therefore, I have determined that the program will select files that do not produce an output when the readPDF package is utilized. Then the chosen files were, as indicated above, divided into .png files and transformed into text; all of this was done by the code depicted in Figure 9 below.

*Figure 9: R code to perform OCR*

```
OCR_rest <- function(){
  for (i in 1:length(corp)){
    if (as.character(nchar(content(corp[[i]])[1])) == 0){
      file =
      paste("../Data/Trajnostna_porocila_raw/",strsplit(NLP::meta(
corp[[i],tag='id'),".pdf",fixed =
TRUE)[[1]][1],".pdf",sep="")
      pngphoto <- pdftools::pdf_convert(file, dpi = 700)
      besedilo <- tesseract::ocr(pngphoto)
      content(corp[[i]]) <- besedilo
    }
  }
  corp <- corp
  corp_dirty <- corp
}
```

*Source: Own work.*

### 2.2.1.2 Cleaning the data

After importing the text into R, one of the most crucial procedures in data science, data cleansing, was performed. All PDF material was saved in an object of the corpus type. The texts were initially preserved by paragraphs, which, the OCR technique described above, does not typically divide precisely. As the paragraphs are not well-defined, I opted to combine them into one and evaluate them by sentence. Then, text was stripped of white space, special characters, and punctuation, with the exception of dots, to preserve the sentence structure. All the Slovenian special characters were converted to SCZ, and all the words were converted to lowercase. The initial processes converted the text into a format that could be lemmatized and examined more quickly in the future. All of this was accomplished via the code provided in Figure 10 below.

Figure 10: R code to perform the corpus cleaning

```
corpus_cleaner <- function(){
  for (i in 1:length(corp)){
    test <-content(corp[[i]])
    test <-paste(test, collapse = '')
    test <-str_squish(test)
    test<-gsub("\\\\.", "\\< JANKOLENC \\>", test)
    test<-gsub("[[:punct:]]+", "", test)
    test<-gsub("\\<JANKOLENC\\>", "\\.", test)
    test<-str_squish(test)
    test<-test[!(test=="")]
    test<-tolower(test)
    Encoding(test) <- "UTF-16"
    write.table(test, file = gsub(" ",
    "",paste("../Data/Trajnostna_porocila_clean/",strsplit(NLP::meta(c
    orp[[i]],tag='id'),"."),fixed =
    TRUE)[[1]][1],"_clean.txt",sep=""), sep = "\t",
    row.names = TRUE, col.names = NA)}
  }
```

Source: Own work.

### 2.2.1.3 Lemmatization

Before discussing technological implementation, it is reasonable to first define lemmatization from a theoretical standpoint. “Lemmatization is a linguistic term that means grouping together words with the same root or lemma, but with different inflections or derivatives of meaning, so they can be analysed as one item. The aim is to remove inflectional suffixes and prefixes, to bring out the word’s dictionary form.” (Techslang, 2019).

I have used a Python-implemented tool called lemmagen for the lemmatization of the cleaned texts. The software was initially developed in a C++ environment by Jernej Virag, Domen Grabec, and Gapšer Žejn of Inštitut Jožef Štefan. The lemmatizer was later converted into a Python module, which I used. I demonstrate how the software lemmatises terms in Table 3 and the following code sample in Figure 11

Table 3: Example of the lemmatization with lemmagen

Original word	Lemmatized word
zaposlen	zaposlen
zaposleni	zaposlen
zaposlena	zaposlen
zaposlenih	zaposlen

Source: GitHub (2019–2020).

Figure 11: Lemmatization example in Python

```
import lemmagen. lemmatizer
from lemmagen. lemmatizer import Lemmatizer

lemmatizer =
Lemmatizer(dictionary=lemmagen.DICTIONARY_SLOVENE)
print(lemmatizer.lemmatize("zaposlenih"))
```

Source: GitHub (2019–2020).

#### 2.2.1.4 AFINN

Text sentiment is determined using the AFINN table. It is a table including over 3000 English words and values ranging from -5 to 5 that indicate the words' emotions. The Danish academic Finn Arup Nielsen produced the AFINN table, which was then translated into Slovene by Rok Martinc (2013). The table was augmented with some sustainability-related terms that have often occurred in the texts. The new terms were added by three academic classifiers, who graded their sentiment on the scale of AFINN and added the field of ESG in another column (as seen in Table 4). The overall tone of a text was obtained by averaging the tones of the words, included in the AFINN table. By comparing the words to the table, any improper or unneeded terms in the texts were automatically eliminated from the analysis since the computer excluded words that are not in the table. The procedure is shown using a sample sentence. The results are presented in Table 5, which includes some terms with values in the AFINN table and others with a value of 0 that were not in the AFINN table and were thus removed from the study.

Table 4: Example of the updated AFINN table

Term	AFINN score	ESG field
odpadek	-2	environment
zdravja	2	Society
razvoj	1	governance

Source: Own work.

Table 5: Example of the sentiment calculation

Sample sentence: V podjetju se trudimo pozitivno delovati na lokalno skupnost.					
podjetje	truditi	pozitivno	delovanje	lokalen	skupnost
0	1	2	0	1	1
Total sentiment: 0.83					

Source: Own work.

### 2.2.1.5 Readability

I picked the Flesch–Kincaid test for judging readability as it is the most commonly used test of readability. 1975 was the year of introduction of the author-named assessments, which are still in use today. The evaluation relies on Equation 1.

$$(F - K) = 206,835 - 1,015 \left( \frac{\text{total words}}{\text{total sentences}} \right) - 84,6 \left( \frac{\text{total syllables}}{\text{total words}} \right) \quad (1)$$

Each report's readability is judged using the aforementioned methodology. The following Table 6 may then be used to classify the findings into reading difficulty categories (Wikipedia, 2022c).

*Table 6: Reading difficulty*

Score	School level (US)	Notes
100,00 > 90,00	5th grade	Very easy to read.
90,0 > 80,0	6th grade	Easy to read.
80,0 > 70,0	7th grade	Fairly easy to read.
70,0 > 60,0	8th & 9th grade	Easily understood by 13- to 15-year-old students.
60,0 > 50,0	10th to 12th grade	Fairly difficult to read.
50,0 > 30,0	College	Difficult to read.
30,0 > 10,0	College Graduate	Best understood by university graduates.
10.0 > 0	Professional	Extremely difficult to read. Best understood by university graduates.

*Source: Wikipedia (2022c).*

Providing samples for further comparison, the standard grade six students' written project (age 12) has a readability index of 60–70 (with a reading grade level of six to seven), whereas the Harvard Law Review has an overall readability score in the low 30s. Harry Potter books have scores between 65 and 82 (Wikipedia, 2022d).

### 2.2.1.6 TF-IDF

TF-IDF is an acronym for the phrase frequency-inverse document frequency. It is an approach that measures the importance of words to specific documents in a group of documents – in our case, in a corpus-type object. As the name suggests, the statistic comprises two parts: term frequency (TF) and inverse document frequency (IDF). The TF-IDF value grows proportionately with the frequency of a word's occurrence in the document and is counterbalanced by the number of corpus documents that include the term (Wikipedia, 2022b).

The first part explains how often a word appears in a particular document and is calculated by dividing a count of word occurrences in a document by the total word appearances in the corpus.

The second part of inverse document frequency measures how frequent or uncommon a word is in all corpus documents. It is the logarithmically scaled inverse fraction of documents including the expression calculated by Equation 2.

$$idf(term) = \ln\left(\frac{\text{number of documents}}{\text{number of documents containing the term}}\right) \quad (2)$$

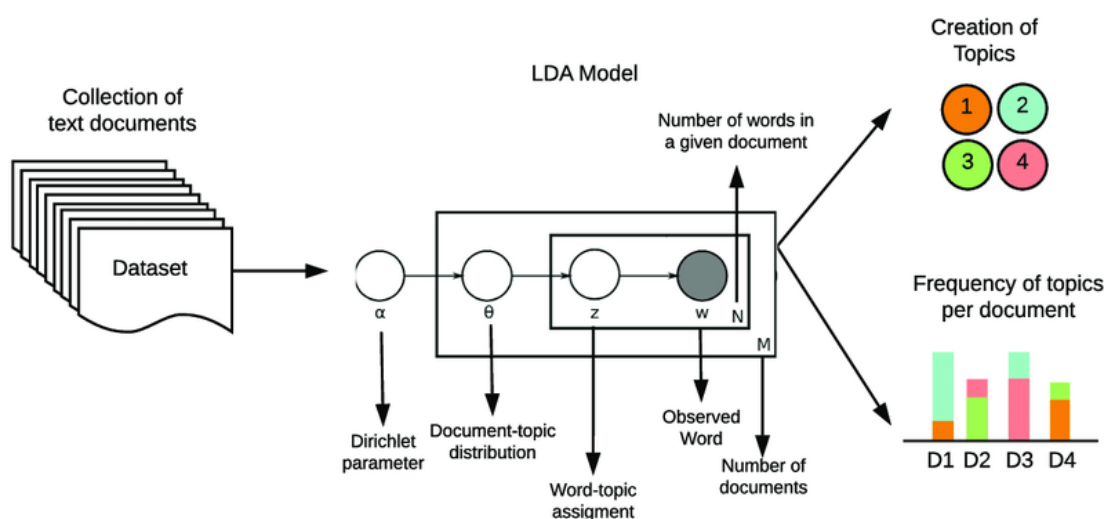
To calculate the TF-IDF score, we need to multiply each part together, namely term frequency and inverse document frequency. The higher the term's score, the more specific the word is to a particular document. In our case, we used TF-IDF to eliminate words specific to certain reports, to avoid overvaluation of reports (Silge & Robinson, 2022).

#### 2.2.1.7 LDA

LDA is an abbreviation for Latent Dirichlet allocation. It is a statistical model for identifying topics inside a text or collection of documents. It is most often used for challenges involving natural language processing. In our situation, we used it to determine what was often addressed in the reports.

Latent Dirichlet allocation assumes that a document comprises a random assortment of latent topics. Topics, that are a mixture of singular words, combine texts saved in the corpus with each topic, defining the probability that the word in question would occur in that topic. LDA speculates that each of the  $M$  documents containing  $N$  words may be represented as a probabilistic Dirichlet distribution on latent themes. Where  $\alpha$  is the Dirichlet parameter prior weight of the subject in a document,  $Z$  is the assignment of a word to a specific topic, and  $W$  is the observed word in the document  $M$ . The process is depicted in Figure 12 below (Buenano-Fernandez, González Gil & Luján-Mora, 2020).

Figure 12: LDA process



Source: Buenano-Fernandez, González Gil & Luján-Mora (2020).

Apart from the above, LDA also assumes that a text (each of  $M$ ) pertains to at least one topic. Under this assumption, it is vital to accurately choose the variables, for which is the subject distribution per document. If a significant value is chosen, the allocation of the topics will be uniform, but a small value inhibits the inference process from dispersing the percentage of likelihood in specific topics. With the method of trial and error, I have come to the conclusion that the most relevant results in my case were established when modelling for three topics (Buenano-Fernandez, González Gil & Luján-Mora, 2020).

## 2.2.2 Financial performance indicators

It is of the highest importance to pick good financial performance metrics to draw meaningful conclusions about the relationship between a company's financial performance and its sustainability report. All of the metrics are computed for each firm that has submitted a sustainability report and for each filing year. The data is derived from AJPES-managed financial reports.

### 2.2.2.1 Selection of indicators and calculations

For the analysis, the chosen indicators are labelled in Table 7 below.

Table 7: Financial performance indicators

Field	Indicator	Formula
Profitability	ROA	net income / assets
	ROE	net income / equity

Table continues

Table 7: Financial performance indicators (cont.)

Field	Indicator	Formula
<b>Margin</b>	EBIT MARGIN	EBIT / sales revenues
	EBITDA MARGIN	EBITDA / sales revenues
<b>Indebtedness</b>	NET DEBT / EBITDA	(debt – cash) / EBITDA
	NET DEBT / ASSETS	(debt – cash) / assets
	INTEREST COVERAGE RATIO	(EBITDA + interest income) / interest expense
<b>Liquidity and solvency</b>	OPERATING CASH FLOW RATIO	operating cash flow (= EBITDA) / current liabilities
	QUICK RATIO	(current assets – inventories) / current liabilities
	CURRENT RATIO	current assets / current liabilities
	SOLVENCY	long term assets / long-term liabilities

Source: Own work.

To assess profitability, I have selected ROA and ROE, since they are the most commonly used financial performance indicators (Choi, Kwak & Choe, 2010) and have been used in many of the recent research papers, as described in section 2. Return on assets (ROA) is a financial metric that assesses the profitability of a business, concerning its total assets (Hargrave, 2022a) ROE is computed by dividing net income by shareholders equity. (Fernando, 2022). ROA incorporates leverage and debt, but ROE does not (Furhmann, 2022).

The authors of the articles (Ching, Gerab & Toste, 2017; Girón, Kazemikhasragh, Cicchiello & Panetti, 2021) have used profit margin to evaluate company performance in their articles which translates into using EBIT margin in this research. I have also used the EBITDA margin to consider depreciation and amortisation.

To define a company's ability to repay its obligations, the primary emphasis is on three ratios: Debt-to-EBITDA, which indicates how many years it would take, if net debt and EBITDA stayed unchanged (Kelton, 2020). Next is the Net Debt to Assets Ratio, which measures the company's leverage. It indicates to what proportion the company's assets are funded by Net Debt (Stockopedia, n.d.). Net Debt was calculated as obligations to funding sources - cash. The interest coverage ratio is a financial metric that is used to evaluate a company's ability to pay the interest on its outstanding debt. It is a measure of the company's debt and profitability and is commonly used by lenders, investors, and creditors to assess risk and determine the feasibility of future borrowing (Hayes, 2022).

The last group of indicators pertains to liquidity and solvency. The operating cash flow ratio assists in determining a company's short-term liquidity and measures the ease, with which current obligations are paid by a company's operating cash flows (Hargrave, 2022b). The

current ratio and quick ratio evaluate a company's capacity to produce sufficient funds to meet all obligations, if they were due to fall simultaneously. The quick ratio is regarded as more conservative than its counterpart, since it does not exclude inventory from its computation (Rhinehart, 2022). The ratio of fixed assets to long-term obligations has been used to determine the solvency of any organization. Solvency is an organization's capacity to pay its obligations using its assets (Chegg, n.d.).

### 2.2.2.2 Descriptive statistics of indicators

In Table 8, I have calculated descriptive statistics of the financial indicators of firms, included in the studied sample.

*Table 8: Financial indicators descriptive statistics*

	<i>Min.</i>	<i>1st.Qu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rd.Qu.</i>	<i>Max.</i>
<i>ROA</i>	2.096E-05	0.017525	0.038205	0.054875	0.078567	0.354567
<i>ROE</i>	6.69907E-05	0.033265	0.078639	0.119402	0.149513	1.254822
<i>EBIT</i>	1,019.06	2,246,466	6,103,084	14,540,337	11,963,426	3.4E+08
<i>EBITDA</i>	81,726.98	6,497,640	14,644,041	30,816,191	29,847,175	4.41E+08
<i>EBIT_MARGIN</i>	0.000200116	0.021859	0.046092	0.064272	0.085146	0.606721
<i>EBITDA_MARGIN</i>	0.008122133	0.070375	0.107237	0.145531	0.174172	1.283795
<i>NET_DEBT_EBITDA</i>	0.253850227	5.933139	8.877433	9.807537	12.21901	45.97898
<i>NET_DEBT_assets</i>	0.428629396	0.941905	0.984356	0.951227	0.996862	0.999981
<i>INTEREST_COVERAGE_RATIO</i>	2.54905752	9.46354	21.59644	38,499.81	73.48387	8,888,566
<i>OPERATING_CASH_FLOW_RATIO</i>	0.024176232	0.26979	0.454199	0.648633	0.862161	5.642072
<i>QUICK_RATIO</i>	0.060639023	0.583925	0.959306	1.238832	1.528054	6.845353
<i>CURRENT_RATIO</i>	0.067430635	0.948726	1.418505	1.730978	2.117359	9.298102
<i>SOLVENCY</i>	0.841026787	2.454075	4.163421	175.6361	11.1859	18599.59

*Source: Own work.*

## 3 ANALYSIS AND RESULTS

In analysis and results I use the classification of the industries presented in Table 9.

*Table 9: Industry classification*

1	Agriculture, forestry and fishing
2	Mining and quarrying
3	Manufacturing
4	Electricity, gas, steam and air conditioning supply
5	Water supply, sewerage, waste management and remediation activities
6	Construction
7	Wholesale and retail trade, repair of motor vehicles and motorcycles
8	Transportation and storage
9	Accommodation and food service activities

*Table continues*



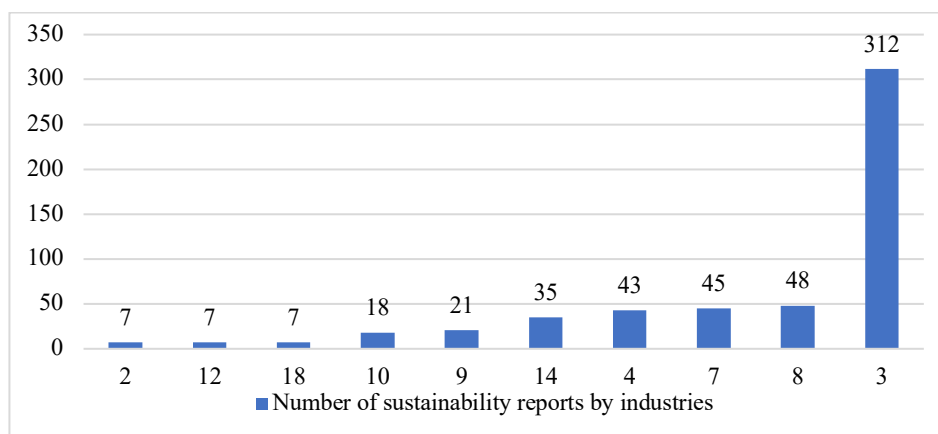
Table 9: Industry classification (cont.)

10	Information and communication
11	Financial and insurance activities
12	Real estate activities
13	Professional, scientific and technical activities
14	Administrative and support service activities
15	Public administration and defence, compulsory social security
16	Education
17	Human health and social work activities
18	Arts, entertainment and recreation
19	Other service activities
20	Activities of households as employers, undifferentiated goods- and services-producing activities of households for own use
21	Activities of extraterritorial organisations and bodies

Source: Statistični urad Republike Slovenije (n.d.).

Firstly, to better grasp the data and its properties, I have analysed some of its fundamental characteristics. One of the facts to check about the dataset was the distribution of sustainability reports between the industries as classified by SKD. Most of the reports, precisely 312, were attributed to firms in the industry 3. The remaining reports were spread evenly between other industries, except 2, 12 and 18, which were underrepresented, with only seven reports in each category as visible in Figure 13.

Figure 13: Number of sustainability reports by industries

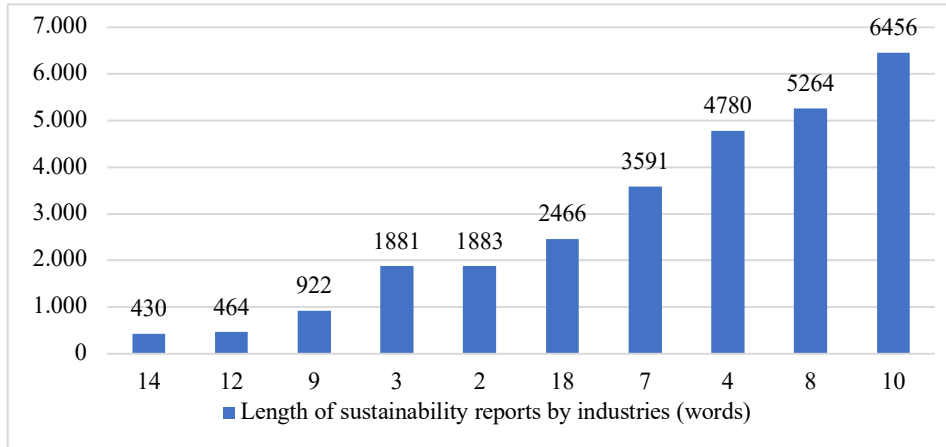


Source: Own work.

The quantity of reports provides little insight into their substance. The first statistic that springs to mind is each report's length. This is depicted in Figure 14. Industry 10 firms provided the average lengthiest reports. The average length of their sustainability reports exceeded 6000 words. The average length of the whole dataset was 2562 words, mainly

owing to the large number of reports in industry 3, that averaged less than 2000 words in each report.

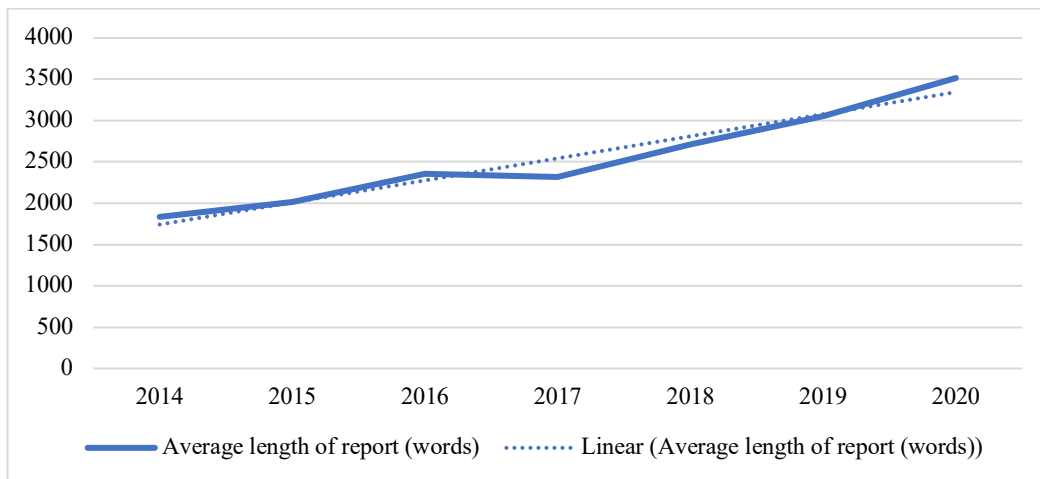
*Figure 14: Average length of sustainability reports by industries*



*Source: Own work.*

As evidence, that sustainability reporting plays a more significant role in business, the accompanying Figure 15 illustrates the average length of sustainability reports over time. From 2014, when the average length of a report was less than 2000 words, to 2020, when the average length of a report exceeded 3,500 words, we can detect an upward trend. This suggests that companies are trying harder each year to provide more comprehensive sustainability reports.

*Figure 15: Average length of sustainability reports through years (words)*

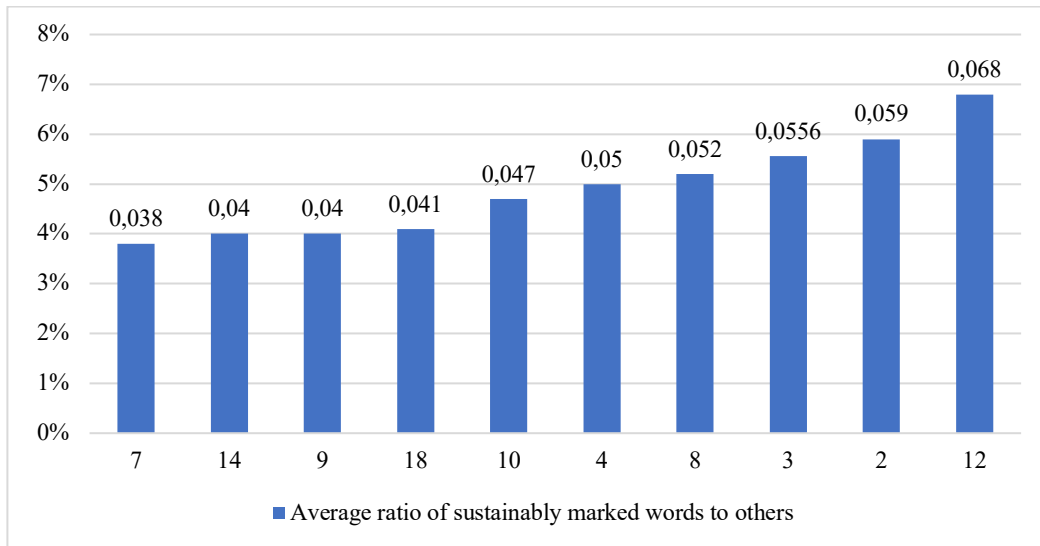


*Source: Own work.*

To go further into the content of the reports, I have created Figure 16 comparing the percentage of sustainability-related terms to other words in each report. The word list is identical to the one presented in the AFFIN table. On average, 5.14 per cent of the words in

each report are rated as sustainable. Due to the small sample sizes, the firm reports in industry 12 are the most above average, and those in industry 14 are among the lowest. This may be due to outliers, since they have more significant effects in smaller samples.

Figure 16: Average ratio of sustainably marked words to others by industries

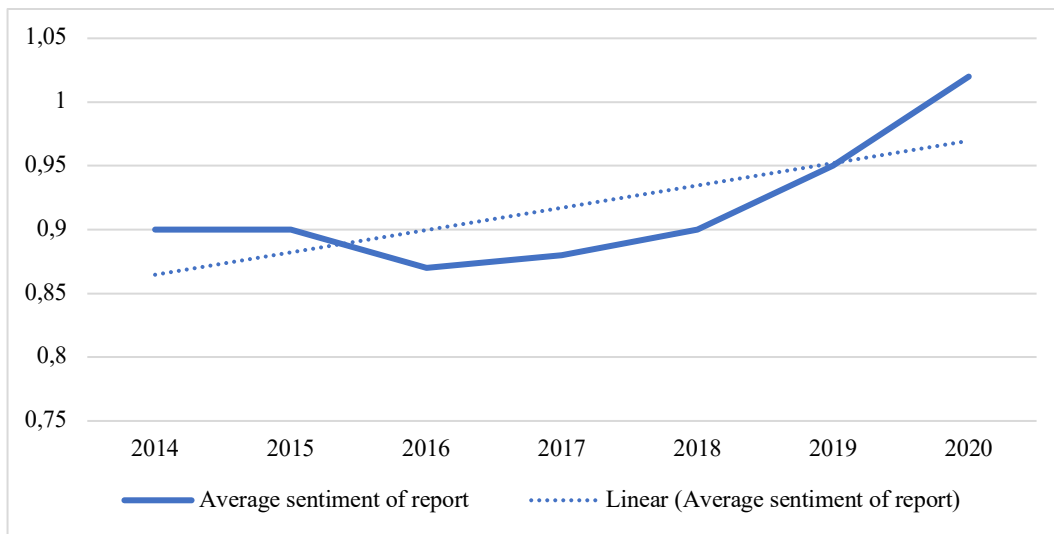


Source: Own work.

### 3.1 AFINN

Moving on from the basic generic properties of sustainability reports to more advanced methods, the sentiment of each report was calculated using the AFINN table and the method explained in chapter 0. To see how sentiment has changed through the years, I have prepared the following figure (Figure 17). Regardless of the industries the companies were in, the sentiment was, on average, experiencing an upwards trend from 0.9 to just above 1. As a reminder, the sentiment of a word can be marked from -5 to 5, so the average sentiment of 1 shows a slightly positive general sentiment.

Figure 17: Average sentiment of sustainability reports through the years

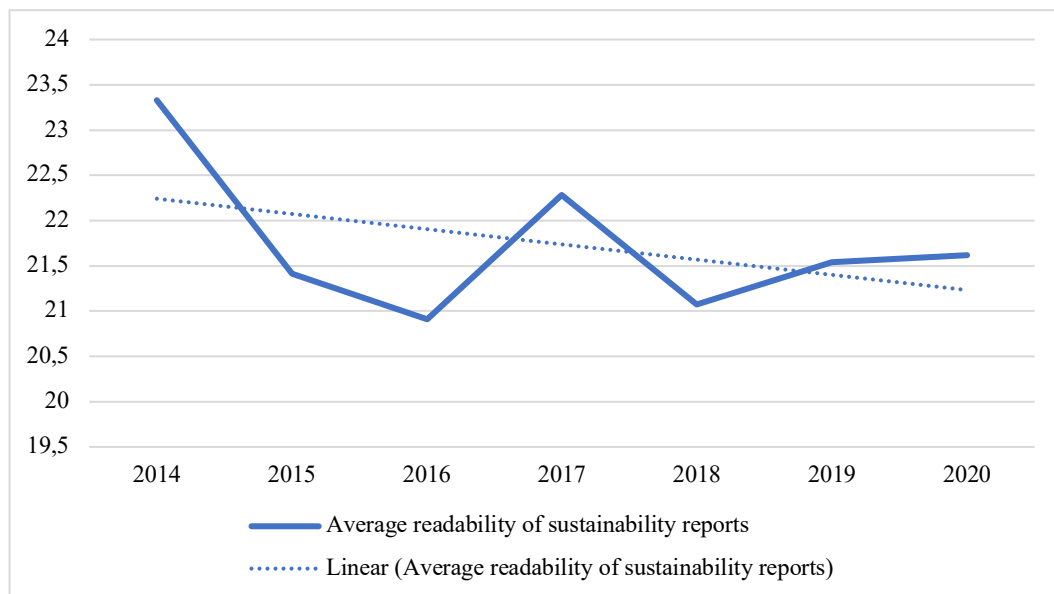


Source: Own work.

### 3.2 Readability

The next metric to evaluate the reports is readability. The calculation was conducted as explained in chapter 2.2.1.5. Average readability was between 21 and 23.5 and showed no apparent pattern, as seen in the Figure 18 below. As per Table 6, the sustainability reporting texts are best understood by graduate students.

Figure 18: Average readability of sustainability reports through the years

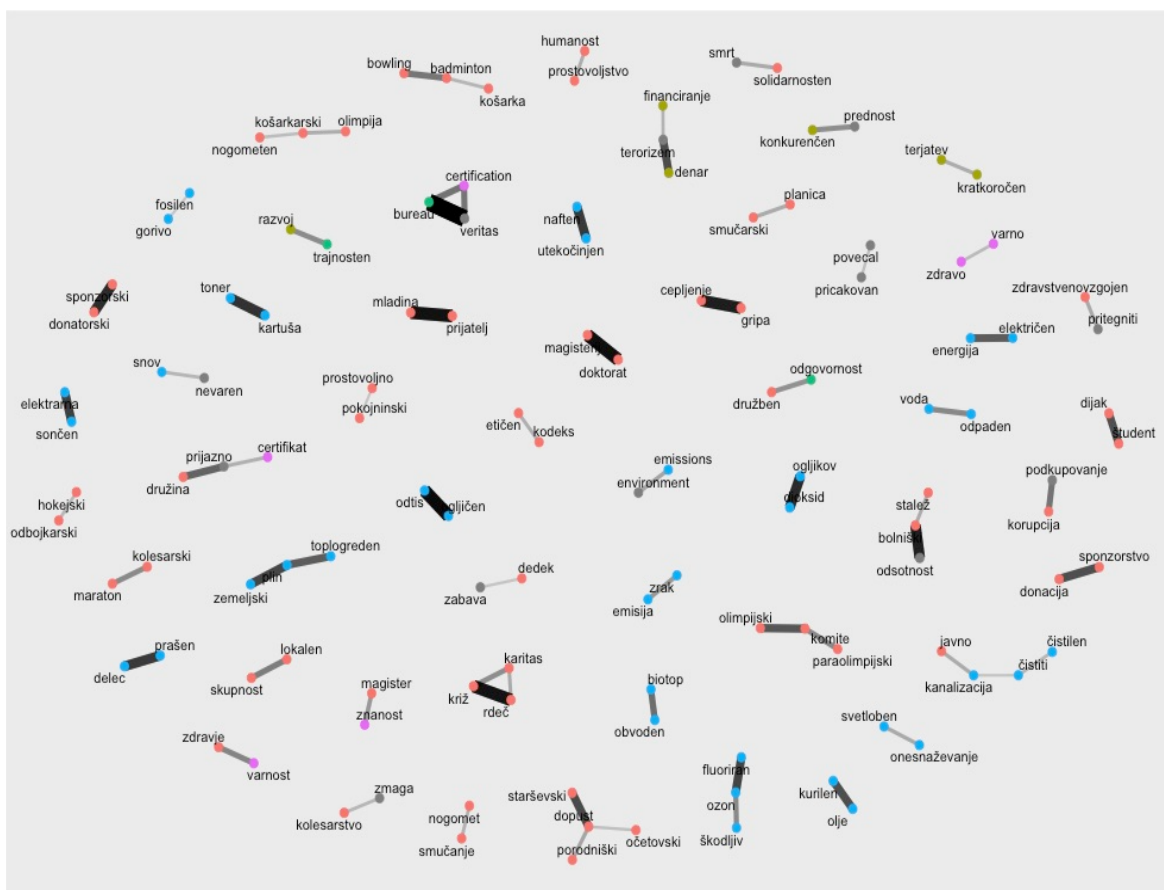


Source: Own work.

### 3.3 Correlation analysis

I conducted a correlation study to discover which kind of word clusters exist in the reports. In the graph below, the strongly associated terms have been connected the concept is depicted in Figure 19. The relationship between two words becomes thicker and less transparent as the correlation between them increases. Words are coloured to represent sustainability quantifiers. We may note that the majority of the terms in each cluster belong to the same sustainability categorization. The cluster "starevski-dopust-porodniški-očetovski" (parental-vacation-maternity leave-paternal) pertains to the social sector of sustainability and depicts paternity leave. Another cluster is "financiranje-terorizem-denar" (financing-terrorism-money), which depicts the problem of funding terrorism and comes under the economic category. "Javno-kanalizacija-čistiti-čistilen" (public-sewer-to clean-cleaning), which relates to the subject of sewage waste management, is a cluster related to the environmental dimension. From this, we can also deduce some general topics that appear in the sustainability report, but this is done with greater precision and detail below.

Figure 19: Clusters of words representing topics



- družbena
- ekonomska, okoljska in družbena
- okoljska
- okoljska in družbena

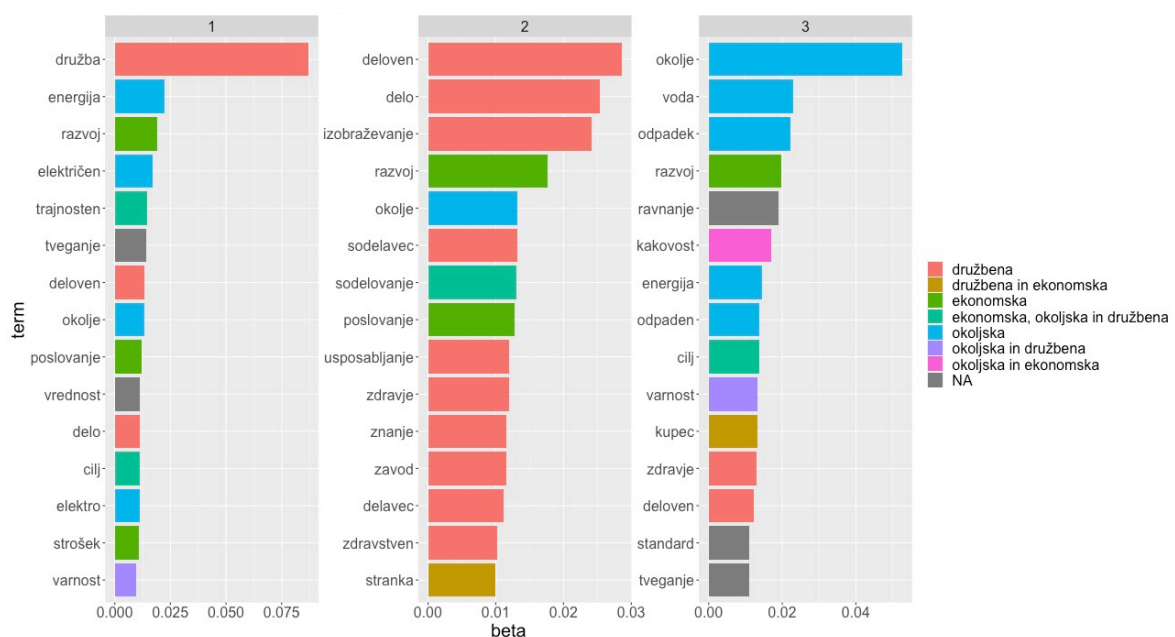
Source: Own work.

### 3.4 LDA

I will investigate the outcomes of the reports' topic analyses in this part. The following figures are generated using the approach described in section 3.2.1.7. As described before, Latent Dirichlet analysis (LDA) is a technique for identifying words that co-occur in distinct topics within a corpus. Each term is given a beta that reflects the per-topic-per-word probability. Thus, a higher beta for term X indicates a greater likelihood that it will occur in topic Y. Figure 20 illustrates that the likelihood of the phrase "okolje" (environment) occurring in topic 1 is around 0.0125; in topic 2 it is approximately 0.012, and in topic 3 it is approximately 0.055. The same idea applies to each figure and phrase in this chapter.

Figure 20 apart from the above shows the outcomes of the LDA algorithm applied to the whole dataset. There are two distinct topics (2 and 3) and one blended topic (1). Except for "družba" (society), all ESG pillars are featured in almost similar proportions in the first mixed topic. Only three of the fifteen words indicating the second issue are unrelated to society, allowing it to be easily distinguished as referring to society. The phrases in subject 2 indicate that it is primarily concerned with growth and knowledge, but also with employees and health. The third subject is similarly categorized, in this instance, as an environmental subject. The words generally describe the environment in terms of ecology, with "okolje" (environment), "voda" (water) and "odpadek" (waste), being the three most prevalent phrases. There are additional references to standards and quality, which may be used in conjunction with any of the phrases in the issue and indicate the environmental awareness of the businesses.

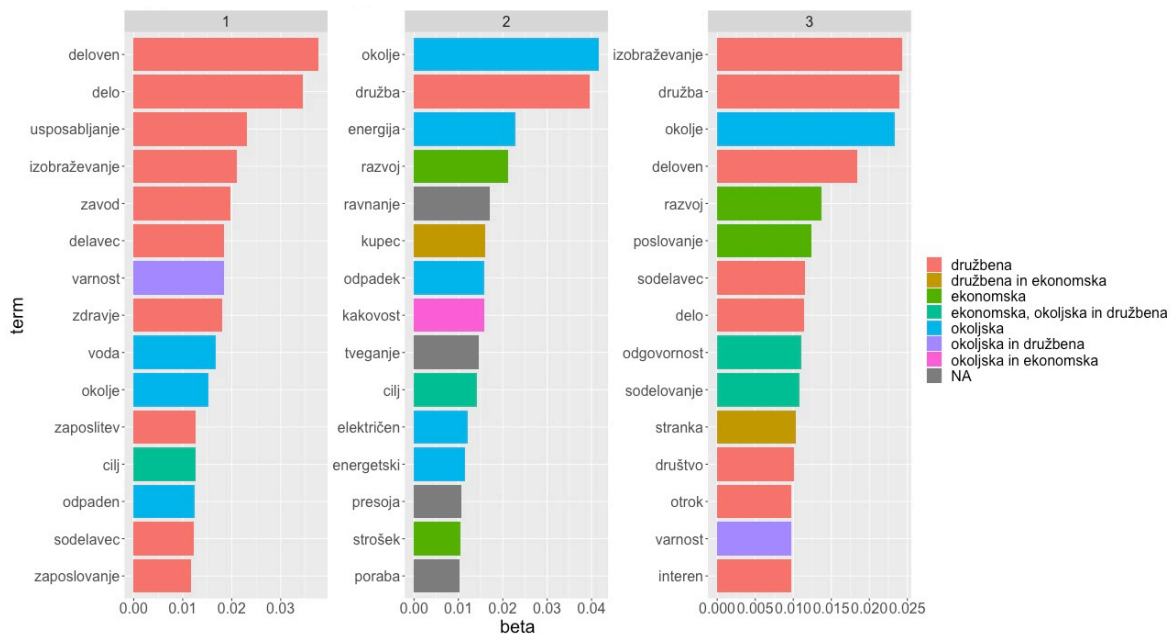
Figure 20: LDA analysis predicting 3 themes on the whole dataset



Source: Own work.

The following figures (Figure 21 and Figure 22) illustrate the variations in subjects between the first and final year in the dataset utilized, independent of the industry, to which the firms belong. In the first picture (Figure 21) are the 2014 LDA analysis findings. Two themes may be categorized as societal and one being quite diverse. In the issues pertaining to society (1 and 3), the majority of the words in the reports pertain to employee training and health, as well as the general themes pertaining to workers. The mixed subject (2) contains terms relevant to all three ESG pillars.

Figure 21: LDA analysis predicting 3 themes in the year 2014 for all companies

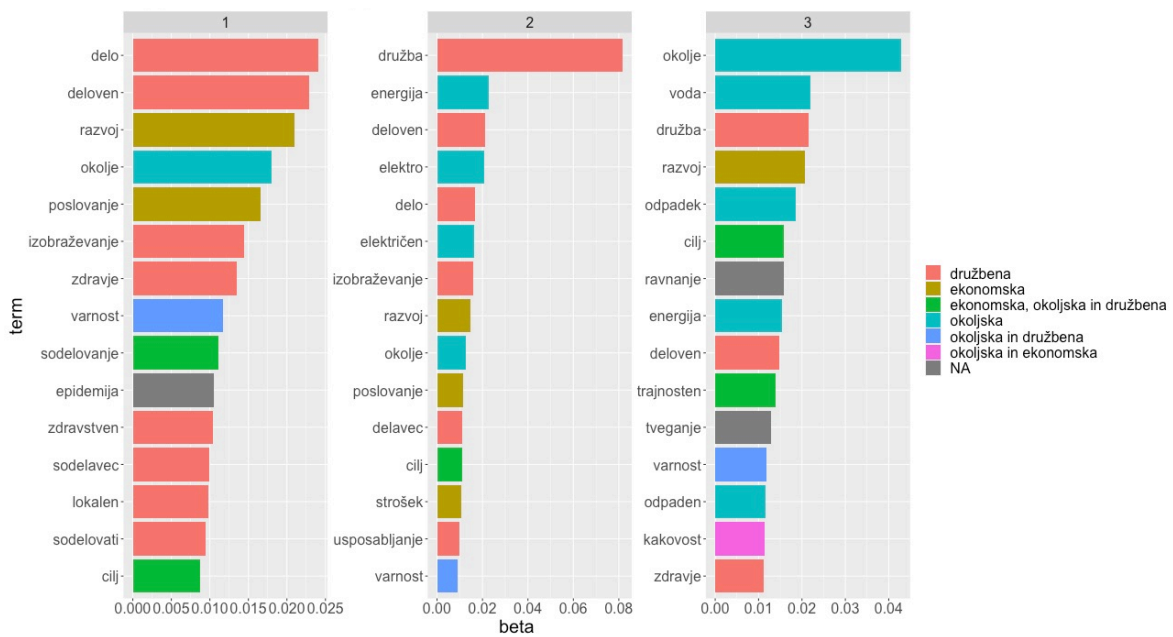


Source: Own work.

The subjects uncovered by LDA in the 2020 reports are shown in Figure 22. In comparison to 2014, there are discernible differences in topic creation. The graph examining the most recent reports shows, that in addition to the social pillar, which dominated in 2014, there is also a prominent environmental subject. As previously mentioned, subject 1 focuses on the company's social influence on the growth and welfare of its workers and now also the surrounding community while shedding light on the COVID-19 epidemic. The second subject is, once again, quite diverse, but it is mainly concerned with electricity, energy, and prices. Environment, which focuses on water and waste disposal, is the last issue.

From this, we can observe a shift in focus regarding the substance of the reports through the years. In 2014, we could observe the focus being mainly on the employees of the firm, in contrast to 2020, where companies are also concerned about the local communities they have entered and the environmental changes they are contributing to.

Figure 22: LDA analysis predicting 3 themes in the year 2020 for all companies



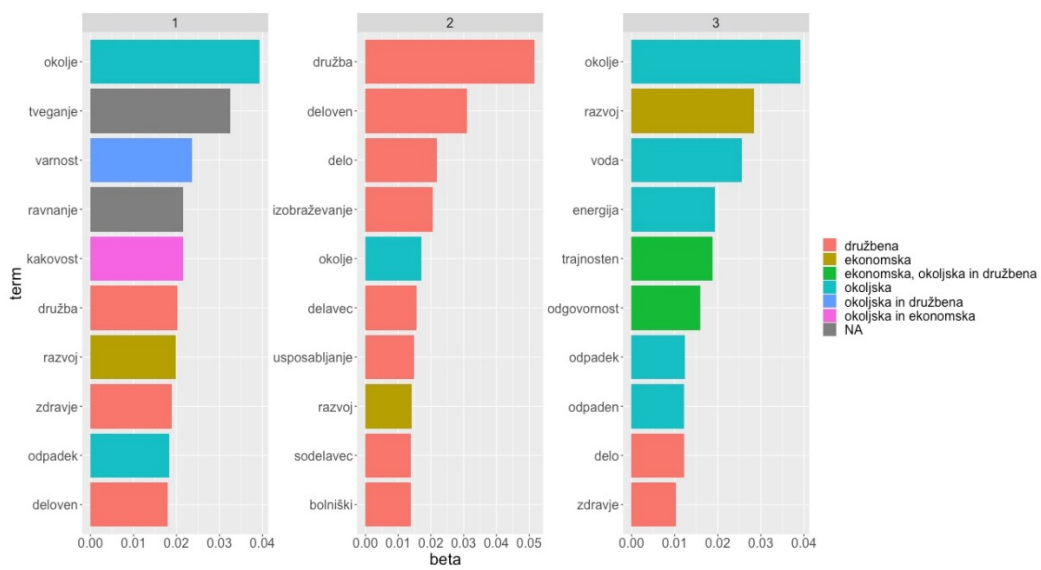
Source: Own work.

In the following figures (Figure 23, Figure 24) I am comparing the results of the LDA analysis of companies in industries (3, 9). For industry 3, we can observe a mixed topic, (1) with no particular focus on any field. Topic 2 mainly consisted of words that are classified as social and mostly concerning the general public and employee training and knowledge. The last topic concerns the environment; more precisely, the topic focuses on responsibility towards energy, water, nature and waste.

In comparison, the firms from industry 9, have similarly to companies from industry 3 a topic concerning society, with very similar terms consisting the topic. In addition, they also have a topic concerning the economic part, mentioning mainly economic growth. There, as one could expect, there is no topic concerning the environment, like in the reports from companies in industry 3; this is due to the fact that industry 9, pertaining to tourism, does not have a particular focus on handling waste and the environment.

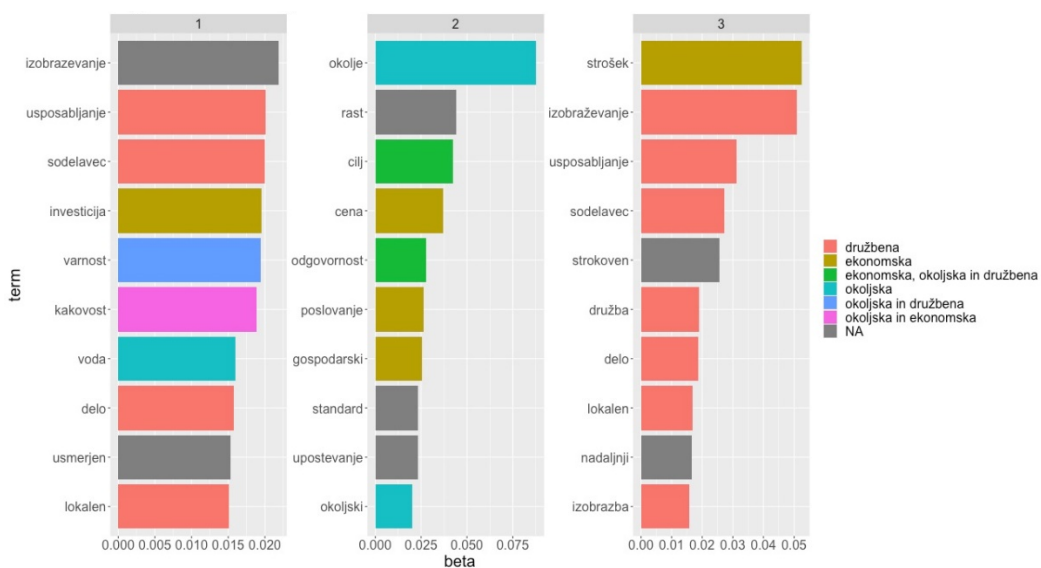


Figure 23: LDA analysis predicting 3 themes in industry 3 for all years



Source: Own work.

Figure 24: LDA analysis predicting 3 themes in industry 9 for all years



Source: Own work.

### 3.5 Regression models

For assessing the relations between the company's financial performance, I have used the following linear regression models presented in Equation 3.

For each  $y_i \in [ROA, ROE, EBIT, EBITDA, EBIT\_MARGIN, EBITDA\_MARGIN, NET\_DEBT, NET\_DEBT\_EBITDA, NET\_DEBT\_assets, INTEREST\_COVERAGE\_RATIO, OPERATING\_CASH\_FLOW\_RATIO, QUICK\_RATIO, CURRENT\_RATIO, SOLVENCY];$

$$y_i = \beta_0 + \beta_1 sentiment + \beta_2 readability + \beta_3 length + \beta_4 sustainable\_to\_others + \sum_{j=5}^{14} \beta_j * z_j + \beta_{15} * velikost \quad (3)$$

Where  $z_j$  is the  $j - 4$ th element of the following vector: [SKD. združeno(2), SKD. združeno(3), SKD. združeno(4), SKD. združeno(7), SKD. združeno(8), SKD. združeno(9), SKD. združeno(12), SKD. združeno(14), SKD. združeno(18)]

Where each of the vector elements is a binary variable, indicating whether the company is in a given industry sector.

The results of each of the models are presented in subchapter 3.7.

### 3.6 Regression assumptions

To find the link between sustainability reporting and the financial performance of the companies I have decided to use linear regression. The model has the following assumptions that should not be violated:

1. Independence of observations
2. No hidden or missing variables
3. Linear relationships
4. Normality of residuals
5. No or little multicollinearity
6. Homoscedasticity
7. No correlation between independent variables and error terms
8. No correlation between the error terms

Each of the above-mentioned assumptions is checked in the following subchapters. All of the assumptions were checked in the manner presented by (Korstanje, 2021).

#### 3.6.1 Independence of observations

This assumption was satisfied by construction, since each report is examined only once (Heidel, n.d.).

### 3.6.2 No Hidden or Missing Variables

The assumption presumes that the model is not mis specified, which was also satisfied by construction.

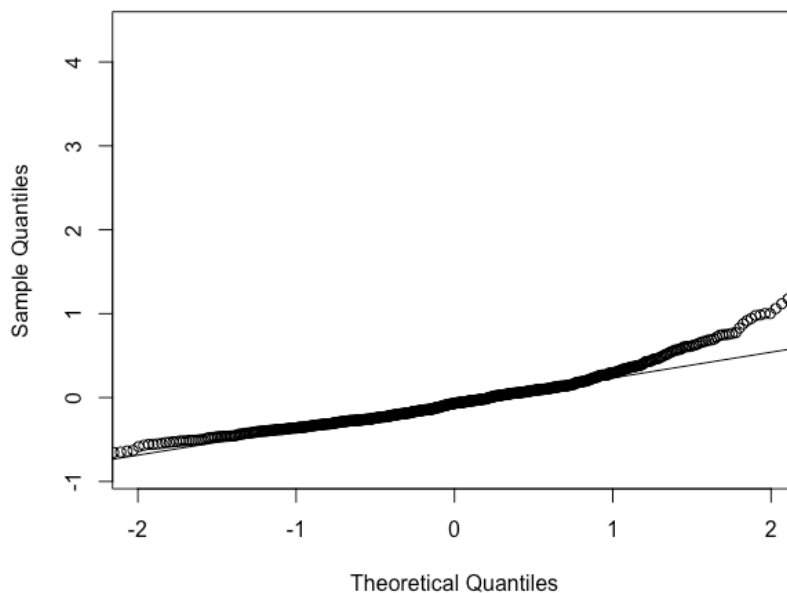
### 3.6.3 Linear relationship

The third assumption presumes linear relations between dependent and independent variables. This was checked using simple scatter plots. The assumption was confirmed after the inspection of the plots.

### 3.6.4 Normality of the residuals

The fourth assumption is tested by using a quantile-quantile plot which is defined as: »A Q-Q plot is a scatterplot created by plotting two sets of quantiles against one another. If both sets of quantiles came from the same distribution, we should see the points forming a line that's roughly straight.« (University of Virginia Library, n.d.) In the case presented in Figure 25 we can observe a minor deviation from the straight line, which indicates a possible violation of the normality presumption of the model. This may be overlooked since the assumption of normality only becomes problematic for small sample sizes. Regarding the central limit theorem, the normality is not as crucial in larger sample sizes such as ours.

*Figure 25: Normal Q-Q plot*



*Source: Own work.*

### 3.6.5 No or little multicollinearity

The fifth assumption of linear regression is that multicollinearity is absent or minimal. Multicollinearity is the phenomenon that occurs when several explanatory variables exhibit high correlations, in that it would not be possible for the model to determine which of the two variables is accountable for a variation of the dependent variable. The VIF reflects the degree of correlation between an independent variable and other independent variables. VIF begins at 1 and has no maximum value. A VIF of 1 is optimal since it implies that this variable does not exhibit multicollinearity. A VIF more significant than 5 or 10 suggests an issue with the model's independent variables (Korstanje, 2021). As it can be seen in Table 10, all the VIF factors are lower than 2, meaning that the fifth assumption is satisfied.

*Table 10: VIF for each of the independent variables*

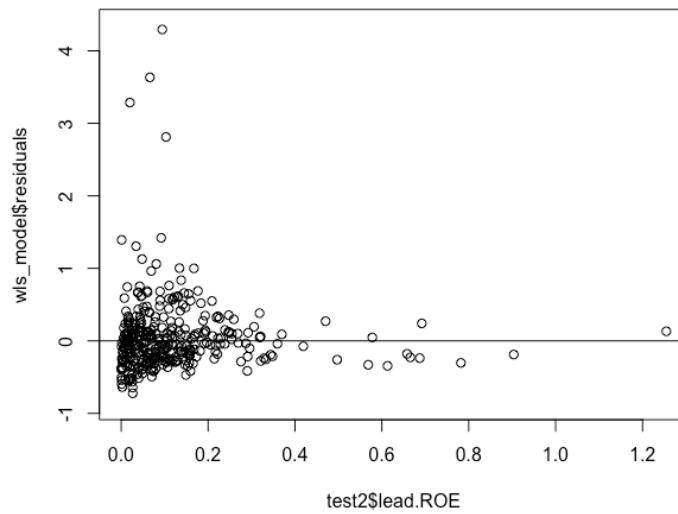
variable	GVIF	DF	GVIF <sup>1/(2*Df)</sup>
sentiment	1.094937	1	1.046393
readability	1.032742	1	1.016239
length	1.207005	1	1.098638
sustainable_to_others	1.194410	1	1.092891
as.factor(Velikost)	1.244405	1	1.115529
as.factor(SKD.združeno)	1.755956	1	1.031773

*Source: Own work.*

### 3.6.6 Homoscedasticity

The sixth assumption concerns homoscedasticity, which is defined as: »a situation in which the error term (that is, the “noise” or random disturbance in the relationship between the independent variables and the dependent variable) is the same across all values of the independent variables« (Statistics Solutions, n.d.). This is easily checked with a scatter plot, comparing the dependent variable and residuals. Looking at Figure 26 we can observe obvious signs of heteroscedasticity (there is no constant deviation from the line at  $y=0$ ). This means a violation of the sixth assumption. This violation is resolved using the weighted least squares model. This is done in a way that each of the observations is assigned a weight that is calculated as the fitted values to the power of negative 2 of a linear model that is regressing the fitted values of the initial model to absolute values of the residuals of the initial model. Consequently, the assumption holds for the new weighted model.

Figure 26: Homoscedasticity plot



Source: Own work.

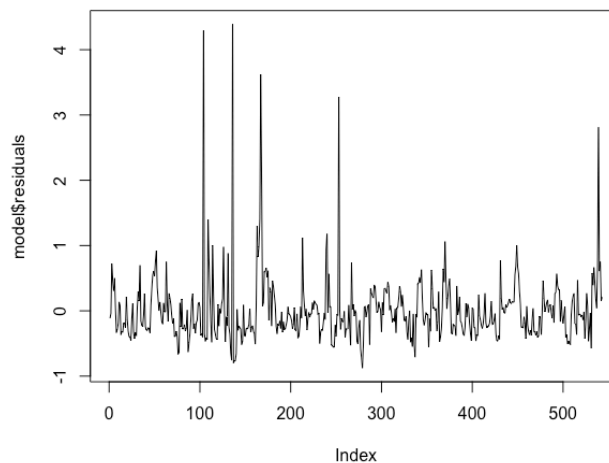
### 3.6.7 All independent variables are uncorrelated with the error term

This was checked in the same manner as in subchapter 3.6.3 but now instead of plotting dependent and independent variables, I have plotted independent variables against residuals. The assumption was confirmed.

### 3.6.8 No correlation between the error terms

To check the last assumption, I have plotted the residuals in Figure 27 in search of any apparent patterns that would lead to a need for an ARMA model. Since no patterns are occurring, the last assumption should hold.

Figure 27: Residual autocorrelation plot



Source: Own work.

After confirming the assumptions, I proceed to provide the results in the next chapter and a discussion in the chapter after the next one.

### 3.7 Regression results

In the following figures I present the results from the regression models used on the dataset.

In Table 11, we can observe the regression results for regressing ROA and ROE. There was no statistical significance of any of the independent variables, except the controls on ROA, while regressing on ROE provided, with three of the independent variables being statistically significant. Sentiment and readability negatively affected ROE, while the report length had a positive impact.

*Table 11: Regression results for financial indicators of PROFITABILITY*

	Dependent variable:	
	lead.ROA	lead.ROE
	(1)	(2)
sentiment	-0.0004 (0.004)	-0.097*** (0.008)
readability	0.0003 (0.0003)	-0.009*** (0.001)
length	-0.00000 (0.00000)	0.00001*** (0.00000)
sustainable to others	-0.033 (0.099)	-0.053 (0.188)
as.factor(Velikost)4	-0.004 (0.007)	-0.073 (0.099)
as.factor(SKD.zdruzeno)12	0.002 (0.013)	0.100 (0.119)
as.factor(SKD.zdruzeno)14	0.015 (0.009)	0.030 (0.057)
as.factor(SKD.zdruzeno)18	0.002 (0.012)	-0.011 (0.066)
as.factor(SKD.zdruzeno)2	0.023 (0.016)	0.045 (0.100)
as.factor(SKD.zdruzeno)3	0.033*** (0.007)	0.074** (0.027)
as.factor(SKD.zdruzeno)4	-0.008 (0.007)	-0.003 (0.022)
as.factor(SKD.zdruzeno)7	0.029*** (0.009)	0.050 (0.045)

*Table continues*

Table 11: Regression results for financial indicators of PROFITABILITY (cont.)

	Dependent variable:	
	lead.ROA	lead.ROE
	(1)	(2)
as.factor(SKD.zdruzeno)8	0.003 (0.007)	-0.042* (0.025)
as.factor(SKD.zdruzeno)9	0.018* (0.010)	0.048 (0.113)
Constant	0.033* (0.012)	0.403** (0.104)
Observations	543	543
R <sup>2</sup>	0.197	0.418
Adjusted R <sup>2</sup>	0.176	0.403

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Source: Own work.

In Table 12 I have depicted regression results for EBIT MARGIN and EBITDA MARGIN. EBIT MARGIN is negatively affected by sentiment, the percentage of sustainable words and positively affected by readability. The latter is positively affected by readability and the percentage of sustainable words, while sentiment was statistically significant in positively affecting the EBITDA margin.

Table 12: Regression results for financial indicators of MARGINS

	Dependent variable:	
	lead.EBIT MARGIN	lead.EBITDA _MARGIN
	(1)	(2)
sentiment	-0.075*** (0.010)	0.041*** (0.004)
readability	0.011*** (0.001)	-0.0002 (0.0003)
length	0.00000 (0.00000)	0.00000 (0.00000)
sustainable to others	-1.075*** (0.118)	0.050 (0.161)
as.factor(Velikost)4	-0.050*** (0.014)	0.010 (0.008)
as.factor(SKD.zdruzeno)12	-0.104** (0.041)	-0.157*** (0.041)
as.factor(SKD.zdruzeno)14	-0.075* (0.043)	-0.138*** (0.041)

Table continues

Table 12: Regression results for financial indicators of MARGINS (cont.)

	Dependent variable:	
	lead.EBIT MARGIN	lead.EBITDA _MARGIN
	(1)	(2)
as.factor(SKD.zdruzeno)18	-0.055 (0.084)	-0.027 (0.061)
as.factor(SKD.zdruzeno)2	0.065 (0.197)	0.112 (0.100)
as.factor(SKD.zdruzeno)3	-0.020 (0.041)	-0.094** (0.040)
as.factor(SKD.zdruzeno)4	-0.006 (0.064)	0.079 (0.052)
as.factor(SKD.zdruzeno)7	-0.008 (0.043)	-0.115*** (0.041)
as.factor(SKD.zdruzeno)8	0.028 (0.069)	0.012 (0.047)
as.factor(SKD.zdruzeno)9	-0.042 (0.059)	-0.070 (0.046)
Constant	0.010 (0.047)	0.165*** (0.042)
Observations	543	543
R <sup>2</sup>	0.454	0.463
Adjusted R <sup>2</sup>	0.439	0.448

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Source: Own work.



Table 13 presents the results of the regressions for NET DEBT EBITDA, NET DEBT ASSETS and INTEREST COVERAGE RATIO as dependent variables. The first shows no statistical significance in any of the independent variables, except the control variables. On the other hand, NET DEBT ASSETS shows statistical significance for the independent variable length and the percentage of sustainable words, while the effect of the first is 0. The regression results of the dependent variable INTEREST COVERAGE RATIO indicate three statistically significant independent variables, namely sentiment, readability and length. All of the significant independent variables affect the dependent one in a positive manner. High regression coefficient in the case of dependent variable INTEREST COVERAGE RATIO come from possible outliers since some companies have zero or no leverage meaning zero to no interest expenses. Meaning that the denominator in the fraction is low and leads to high ratio value. Outlier treatment should be done in future research.

Table 13: Regression results for financial indicators of INDEBTEDNESS

	<i>Dependent variable:</i>		
	lead.NET DEBT EBITDA	lead.NET DEBT assets	lead.INTEREST COVERAGE RATIO
	(1)	(2)	(2)
sentiment	0.070	-0.006	46,744.390***
	(0.482)	(0.006)	(9,838.537)
readability	0.005	0.0001	802.747*
	(0.024)	(0.0001)	(407.387)
length	0.0001	-0.00000**	1.613*
	(0.0001)	(0.00000)	(0.732)
sustainable to others	2.309	0.656**	591,798.700
	(11.932)	(0.085)	(460,548.600)
as.factor(Velikost)4	-0.381	0.046*	8,612.396
	(0.913)	(0.025)	(17,286.070)
as.factor(SKD.zdruzeno)12	7.331***	-0.089	14,790.050
	(2.538)	(0.056)	(23,854.250)
as.factor(SKD.zdruzeno)14	5.380**	-0.020	55,398.890**
	(1.431)	(0.014)	(16,678.240)
as.factor(SKD.zdruzeno)18	-0.956	-0.067	45,003.740
	(1.720)	(0.060)	(42,803.310)
as.factor(SKD.zdruzeno)2	-0.720	0.013	52,214.910
	(1.738)	(0.013)	(77,079.560)
as.factor(SKD.zdruzeno)3	1.554	-0.026**	27,717.690**
	(1.029)	(0.012)	(13,728.600)
as.factor(SKD.zdruzeno)4	4.235***	0.018	58,881.170***
	(1.254)	(0.012)	(13,252.290)
as.factor(SKD.zdruzeno)7	2.762**	0.007	57,032.370**
	(1.208)	(0.014)	(13,998.530)
as.factor(SKD.zdruzeno)8	4.248***	0.002	40,596.180**
	(1.226)	(0.013)	(16,622.450)
as.factor(SKD.zdruzeno)9	1.764	-0.034	89,601.740**
	(1.436)	(0.042)	(16,140.140)
Constant	7.360***	0.901***	-109,504.400***
	(1.584)	(0.029)	(27,632.040)
Observations	543	543	543
R <sup>2</sup>	0.096	0.355	0.769
Adjusted R <sup>2</sup>	0.072	0.338	0.763

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Source: Own work.

The second to last Table 14) of the chapter consists of data from the linear regression models of OPERATING CASH FLOW RATIO, QUCIK RATIO, CURRENT RATIO and SOLVENCY. The variable OPERATING CASH FLOW RATIO is positively affected only by sentiment and some control variables. In the cases of the other two ratios, there were no significant independent variables, only the control ones. On the other hand, with SOLVENCY I have discovered statistically significant variables, namely sentiment, readability and the ratio of sustainable words.

*Table 14: Regression results for financial indicators of LIQUIDITY AND SOLVENCY*

	Dependent variable:			
	lead.OPERATING_CASH FLOW.RATIO	lead.QUICK_ RATIO	lead.CURRENT RATIO	lead.SOLV ENCY
	(1)	(2)	(3)	(4)
sentiment	0.123*** (0.046)	0.082 (0.078)	0.136 (0.101)	-108.957*** (16.486)
readability	0.0001 (0.002)	0.001 (0.005)	-0.0003 (0.006)	4.037*** (1.163)
length	0.00001 (0.00001)	0.00000 (0.00001)	-0.00001 (0.00001)	-0.001 (0.004)
sustainable to others	1.575 (1.253)	-1.355 (2.162)	-2.778 (2.686)	-3,607.476 ** (848.602)
as.factor(Velikost )4	0.108 (0.073)	-0.326* (0.188)	-0.269 (0.176)	37.116 (49.072)
as.factor(SKD.zd ruzeno)12	-0.602** (0.182)	0.115 (0.387)	0.042 (0.375)	-77.860 (92.590)
as.factor(SKD.zd ruzeno)14	-0.363* (0.167)	0.101 (0.266)	-0.048 (0.263)	-87.808 (99.615)
as.factor(SKD.zd ruzeno)18	-0.122 (0.238)	-0.454* (0.263)	-0.610** (0.239)	-14.229 (95.594)
as.factor(SKD.zd ruzeno)2	0.552 (0.380)	-0.435* (0.262)	-0.324 (0.287)	11.022 (90.672)
as.factor(SKD.zd ruzeno)3	-0.189 (0.154)	0.096 (0.212)	0.490** (0.217)	104.683 (89.188)
as.factor(SKD.zd ruzeno)4	0.368* (0.193)	0.102 (0.241)	0.386 (0.260)	-54.025 (135.503)
as.factor(SKD.zd ruzeno)7	-0.221 (0.164)	-0.395* (0.212)	-0.116 (0.226)	-13.595 (89.167)
as.factor(SKD.zd ruzeno)8	-0.036 (0.171)	0.557** (0.269)	0.484* (0.257)	246.546 (342.720)
as.factor(SKD.zd ruzeno)9	-0.097 (0.192)	-0.277 (0.254)	-0.411* (0.246)	-111.097 (100.591)

*Table continues*

Table 14: Regression results for financial indicators of LIQUIDITY AND SOLVENCY  
(cont.)

	Dependent variable:			
	lead.OPERATING_CASH FLOW.RATIO	lead.QUICK_ RATIO	lead.CURRENT RATIO	Lead.SOLV ENCY
	(1)	(2)	(3)	(4)
Constant	0.468** (0.187)	1.448* (0.312)	1.740** (0.320)	247.974** (102.539)
Observations	543	543	543	543
R <sup>2</sup>	0.152	0.108	0.168	0.947
Adjusted R <sup>2</sup>	0.130	0.085	0.146	0.946

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Source: Own work.

Table 15 depicts the results of regressing the company's sector on length and readability as the metrics to measure the quality of the sustainability report; regarding length, all of the sectors except sectors 8 and 4 show statistical significance. Regarding readability, there was no statistical significance in any of the independent variables.

Table 15: Regression results for regressing company sector on quality variables

	Dependent variable:	
	length	readability
	(1)	(2)
as.factor(SKD.zdruzeno)12	-6,001.135*** (1,678.744)	5.140 (3.753)
as.factor(SKD.zdruzeno)14	-6,034.849*** (1,093.119)	-0.002 (2.444)
as.factor(SKD.zdruzeno)18	-3,999.278** (1,678.744)	-1.212 (3.753)
as.factor(SKD.zdruzeno)2	-4.582.563*** (1,678.744)	-0.861 (3.753)
as.factor(SKD.zdruzeno)3	-4,583.925*** (913.573)	-0.053 (2.042)
as.factor(SKD.zdruzeno)4	-1,685.650 (1,058.021)	0.622 (2.365)
as.factor(SKD.zdruzeno)7	-2,874.500*** (1,051.060)	-2.193 (2.350)
as.factor(SKD.zdruzeno)8	-1,200.507 (1,041.633)	-0.269 (2.329)
as.factor(SKD.zdruzeno)9	-5,543.516*** (1,210.560)	- 3.073 (2.706)

Table continues

Table 15: Regression results for regressing company sector on quality variables (cont.)

	Dependent variable:	
	length	readability
	(1)	(2)
Constant	6,465.278*** (888.308)	21.994*** (1.986)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Source: Own work.

All of the above-mentioned results from all the figures will be discussed in detail and given the proposed reasoning in the following section (Chapter 4).

## 4 DISCUSSION

### 4.1 Interpretation of main findings and conclusions

I will start the discussion by answering all of the research questions. For easier comprehension results are summarised in Table 16.

The first question pertains to the structure of the sustainability reports and not to their contents; thus, the variables explaining the structure that I have used are readability and report length. Among different fields of financial indicators, there was no apparent difference. It turned out that in most cases, when the readability of sustainability reports was statistically significant, readability positively affected the financial indicator. This means that the easier the text was to read, the better was the financial performance indicator. The underlying reason for this could be that firms want to address the general population when discussing sustainability. To make such an address companies should use easier, plainer language. When people understand their attempts for sustainability, they might prefer them over others when making purchasing decisions which would lead to better financial performance.

The length of the report was significant only in two cases, when taking into regard ROE and INTEREST COVERAGE RATIO. Nevertheless, in both cases of significance the length positively affected the financial indicator, meaning that more comprehensive reports resulted in better financial indicator values. This might be due to more comprehensive reports, providing more insight into the company and it signals greater transparency.

To answer research question 1, to what extent do a company efficient and well-structured sustainability reports relate to good financial performance? The analysis reveals the following: in general, extensive well-structured reports are related to better financial performance. This is in line with the findings of (Ching, Gerab & Toste, 2017; Al-Wattar,

Almagtome & Al-Shafeay, 2019), who have, apart from focusing on the contents of the reports, focused on quality and have found positive relations between quality and financial performance. This relation seems intuitive in the manner of “better reports better performance” and is confirmed by scientific findings.

The second research question pertains to the relationship between the report contents and company financial performance. For this purpose, I have used sentiment and the percentage of sustainable words as independent variables.

The sentiment was found to be statistically significant for at least one indicator from each of the financial performance fields. However, the relationship was positive in 50% of the cases and negative in 50% as well. Since we expect higher scores on the financial indicators for better financial performance, even if significant, these results do not tell us much about the general case.

Similarly, although the percentage of sustained words was significant in some cases, the direction of the effect was not even approximately the same. Consequently, I cannot conclude the final general effect. This effect could be due to the reason that the AFINN table was not extensive enough or not evaluated in the correct manner.

The last research question concerns the connection between a company sector and its report quality. As mentioned above, the quality was measured using the metrics of readability and length. The results show that a company in a particular industry does not affect the readability of its report. On the other hand, if a company is involved in an industry that is neither 4 or 8, then its report length is affected by the industry it is in. The most comprehensive reports come from industry 10 (information and communication) as this is an industry with one of the highest churn rates<sup>12</sup> (Statista, n.d.). Consequentially the companies of this industry prepare extensive and comprehensive sustainability reports to try to convince their customers of their efforts towards sustainability with the final goal being customer retention.

To conclude and answer the final question: what can be inferred about future financial performance from the content of a sustainability report and its characteristics? From my results it can be inferred that there is a relationship between the content of sustainability reports and the company financial performance, but the way the content affects the financial performance is not known because the results are unclear.

---

<sup>12</sup> »Churn rate, sometimes known as attrition rate, is the rate at which customers stop doing business with a company over a given period of time.« (Salesforce, n.d.)

Table 16: Results Summary

Field	Dependent variable	Sentiment	Readability	Length	% of sustainable words	Adjusted $R^2$
<b>Profitability</b>	ROA	/	/	/	/	0.176
	ROE	-	-	+	/	0.403
<b>Margin</b>	Ebit margin	-	+	/	-	0.439
	Ebitda margin	+	/	/	/	0.448
<b>Indebtedness</b>	Net debt / ebitda	/	/	/	/	0.073
	Net debt / assets	/	/	/	+	0.338
	Interest coverage ratio	+	+	+	/	0.763
<b>Liquidity and solvency</b>	Operating cash flow ratio	+	/	/	/	0.130
	Quick ratio	/	/	/	/	0.085
	Current ratio	/	/	/	/	0.146
	Solvency	-	+	/	-	0.946

*Vir: Own work.*

## 4.2 Limitations and future research

This research brings some insight into how a structure and quality of a sustainability report is important for the financial aspect of the firm. It portrays well the importance of sustainability reports and sets the path for further research, but as all scientific research it has some downsides or obstacles.

During the course of my research, I encountered several obstacles. Let me concentrate on the technical ones first. As previously indicated, some of the reports were not machine-readable, which complicated the procedure. It also made the procedure significantly less precise, since OCR techniques are not 100 per cent exact. This issue should be fixed as soon as the new regulation, requiring companies to produce machine-readable reports, goes into effect.

Another problem I have come upon were embedded reports, which meant that the sustainability reports were incorporated into the company annual report. This necessitated manual analysis of the reports to determine which section of the annual report is occupied by the sustainability report. Occasionally, sustainability reports are dispersed throughout the yearly report. This made it even more challenging to collect all the essential information for the sustainability report in one location. To tackle this problem in the future, it should suffice

to collaborate with an expert in AI or IT to develop a computer that can automatically detect and extract sustainability reports from annual reports.

With the assistance of linguists, increasing the AFINN table is another potential area for development. The table and its scores should be reconsidered and augmented by adding more Slovene terms. This extension could serve as the table for determining the tone of any Slovene-language document, not only sustainability reports.

Future research is necessary to objectivise and further investigate the research topic pertaining to the quality of sustainability reports and financial performance of the companies. While discussing the quality of sustainability reports, in addition to length and readability, more measures should be applied. More metrics might result in more accurate regression models, since most of those used in this research had low adjusted R squared.

## REFERENCE LIST

1. Al-Wattar, Y. M. A., Almagtome, A. H. & Al-Shafeay, K. M. (2019). The role of integrating hotel sustainability reporting practices into an Accounting Information System to enhance Hotel Financial Performance: Evidence from Iraq. *African Journal of Hospitality, Tourism and Leisure*, 8(5), 1–16.
2. Asuquo, D. A. I., Temitayo, D. E. & Raphael, O. U. (2018). The Effect of Sustainability Reporting on Corporate Performance of Selected Quoted Brewery Firms in Nigeria. *International Journal of Business & Law Research*, 6(3), 1–10.
3. Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., Lowe, W., Müller, C. & European Research Council. (2022). *Quanteda: Quantitative Analysis of Textual Data (3.2.1)*. Retrieved 14 September 2022 from <https://CRAN.R-project.org/package=quanteda>
4. Buallay, A., Hamdan, A. & Barone, E. (2019). Sustainability reporting and firm's performance: Comparative study between manufacturing and banking sectors. *International Journal of Productivity and Performance Management*, 69(3), 431–445.
5. Buenano-Fernandez, D., González, M., Gil, D. & Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access*, 8, 318–330.
6. Chegg. (n.d.). *Learn About Ratio Of Fixed Assets To Long-Term Liabilities*. Retrieved 14 September 2022 from <https://www.chegg.com/learn/accounting/accounting/ratio-of-fixed-assets-to-long-term-liabilities>
7. Ching, H. Y., Gerab, F. & Toste, T. H. (2017). The Quality of Sustainability Reports and Corporate Financial Performance: Evidence From Brazilian Listed Companies. *SAGE Open*, 7(2), 2158244017712027.
8. Choi, J. S., Kwak, Y. M. & Choe, C. (2010). Corporate Social Responsibility and Corporate Financial Performance: Evidence from Korea. *Australian Journal of Management*, 35(3), 291–311.



9. Circularise. (2022). *Corporate Sustainability Reporting Directive (CSRD) Explained*. Retrieved 26 August 2022 from <https://www.circularise.com/blog/corporate-sustainability-reporting-directive-csrd-explained>
10. Ermenc, A., Klemenčič, M. & Rejc Buhovac, A. (2017). Sustainability Reporting in Slovenia: Does Sustainability Reporting Impact Financial Performance? In P. Horváth & J. M. Pütter (eds.), *Sustainability Reporting in Central and Eastern European Companies: International Empirical Insights* (pp. 181–197). Cham: Springer International Publishing.
11. ESG – The Report. (n.d.). *What are the 3 components of the triple bottom line approach?* Retrieved 6 September 2022 from <https://www.esgthereport.com/what-is-esg/the-g-in-esg/what-are-the-3-components-of-the-triple-bottom-line-approach>
12. European Commission. (2019). *The European Green Deal sets out how to make Europe the first climate-neutral continent by 2050, boosting the economy, improving people's health and quality of life, caring for nature, and leaving no one behind*. Retrieved 6 September 2022 from [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_19\\_6691](https://ec.europa.eu/commission/presscorner/detail/en/ip_19_6691)
13. European Commission. (n.d.). *EU taxonomy for sustainable activities*. Retrieved 7 September 2022 from [https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities\\_en](https://finance.ec.europa.eu/sustainable-finance/tools-and-standards/eu-taxonomy-sustainable-activities_en)
14. Fernando, J. (2022, June 27). *Return on Equity (ROE) Calculation and What It Means*. Investopedia. Retrieved 14 September 2022 from <https://www.investopedia.com/terms/r/returnnonequity.asp>
15. Fuhmann, R. (2022, March 22). *Return on Equity (ROE) vs. Return on Assets (ROA): What's the Difference*. Retrieved 14 September 2022 from <https://www.investopedia.com/ask/answers/070914/what-are-main-differences-between-return-equity-roe-and-return-assets-roa.asp>
16. Girón, A., Kazemikhasragh, A., Cicchiello, A. F. & Panetti, E. (2021). Sustainability Reporting and Firms' Economic Performance: Evidence from Asia and Africa. *Journal of the Knowledge Economy*, 12(4), 1741–1759.
17. GitHub. (2019–2020). *Slovenian Lemmatizer for Python*. Retrieved 14 September 2022 from <https://github.com/izacus/SlovenianLemmatizer-Python>
18. Global Reporting Initiative. (n.d.). *GRI – Standards*. Retrieved 20 August 2022 from <https://www.globalreporting.org/standards/>
19. Hargrave, M. (2022a, June 14). *Return on Assets (ROA): Formula and 'Good' ROA Defined*. Retrieved 14 September 2022 from <https://www.investopedia.com/terms/r/returnonassets.asp>
20. Hargrave, M. (2022b, Mach 23). *Analyzing a Company's Liquidity Using the Operating Cash Flow Ratio*. Retrieved 14 September 2022 from <https://www.investopedia.com/terms/o/ocfratio.asp>
21. Hayes, A. (2022, June 17) *Interest Coverage Ratio: The Formula, How It Works, an Example*. Retrieved 14 September 2022 from <https://www.investopedia.com/terms/i/interestcoverageratio.asp>

22. Heidel, E. (n.d.). *Independence of Observations Means Each Study Participant is Independent of All Other Observations*. Retrieved 17 September 2022 from <https://www.scalestatistics.com/independence-of-observations.html>
23. Hiller, J. & Herbst-Bayliss, S. (2021, April 15). *Exxon, activist spend over \$65 million in battle for oil giant's future*. Retrieved 7 September 2022 from <https://www.reuters.com/article/us-exxon-mobil-proxy-activist-idCAKBN2C22SO>
24. Howard-Grenville, J. (2021, January 22). ESG Impact Is Hard to Measure—But It's Not Impossible. *Harvard Business Review*. Retrieved 7 November 2022 from <https://hbr.org/2021/01/esg-impact-is-hard-to-measure-but-its-not-impossible>
25. Hussain, N. (2015). *Impact of Sustainability Performance on Financial Performance: An Empirical Study of Global Fortune (NI100) Firms* (SSRN Scholarly Paper ID 2616835). Retrieved 7 November 2022 from <https://papers.ssrn.com/abstract=2616835>
26. Institute for Statistics and Mathematics (n.d.). *Comprehensive R Archive Network – R Project*. Retrieved November 6 2022 from <https://cran.r-project.org/web/packages/tesseract/vignettes/intro.html>
27. Kasbun, N., Teh, B. & Ong, T. (2016). Sustainability reporting and financial performance of Malaysian public listed companies. *Institutions and Economies*, 8, 78–93.
28. Kelton, W. (2020, June 27). *Net Debt-to-EBITDA Ratio*. Retrieved 14 September 2022 from <https://www.investopedia.com/terms/n/net-debt-to-ebitda-ratio.asp>
29. Korstanje, J. (2021, June 20). *Assumptions of linear regression*. Retrieved 14 September 2022, <https://towardsdatascience.com/assumptions-of-linear-regression-fdb71ebeeaa8b>
30. Kumar, A. (2022, April 27). *ESG Metrics & KPIs and ESG Reporting Concepts*. Retrieved 14 September 2022, <https://vitalflux.com/esg-metrics-and-kpis-what-esg-team-needs-to-know>
31. Laskar, N. (2016). Does Sustainability Reporting Enhance Firms Profitability? A Study on Select Companies from India and South Korea. *Indian Journal of Corporate Governance*, 12(1), 2–20.
32. Loh, L., Thomas, T., & Wang, Y. (2017). Sustainability Reporting and Firm Value: Evidence from Singapore-Listed Companies. *Sustainability*, 9(11), 2112.
33. Martinc, R. (2013). *Merjenje sentimenta na družbenem omrežju Twitter : izdelava orodja ter evaluacija* (Master Thesis). Ljubljana: University of Ljubljana, Faculty of Social Sciences.
34. Motwani, S. S. & Pandya, H. B. (2016). Evaluating the impact of sustainability reporting on financial performance of selected Indian companies. *International Journal of Research in IT and Management*, 6(2), 6.
35. Nguyen, T. (2020). An Empirical Study on the Impact of Sustainability Reporting on Firm Value. *Journal of Competitiveness*, 12, 119–135.
36. Pizzi, S., Corbo, L. & Caputo, A. (2021). Fintech and SMEs sustainable business models: Reflections and considerations for a circular economy. *Journal of Cleaner Production*, 281, 125217.

37. PricewaterhouseCoopers. (n.d.). *Sustainable Finance Disclosure Regulation (SFDR)*. Retrieved 7 September 2022 from <https://www.pwc.be/en/challenges/sustainability/sustainability-assurance-and-reporting/sustainable-finance-disclosure-regulation-sfdr.html>
38. Rhinehart, C. (2022, January 12). *How Do the Current Ratio and Quick Ratio Differ?* Retrieved 14 September 2022 from <https://www.investopedia.com/ask/answers/062714/what-are-main-differences-between-current-ratio-and-quick-ratio.asp>
39. Rio ESG. (2021, may 6). *ESG Ratings vs. ESG Reporting: What's the Difference?* Retrieved 26 August 2022 from <https://www.rio.ai/blog/esg-ratings-vs.-esg-reporting-whats-the-difference>
40. S&P Global. (n.d.). *Investor demand for financially material Environmental Social and Governance (ESG) information is growing rapidly*. Retrieved 21 August 2022 from [https://www.spglobal.com/\\_assets/documents/ratings/esg/esg\\_evaluation\\_brochure\\_digital.pdf](https://www.spglobal.com/_assets/documents/ratings/esg/esg_evaluation_brochure_digital.pdf)
41. Salesforce. (n.d.). *How to Calculate Customer Churn Rate and Revenue Churn Rate*. Retrieved 7 November 2022 from <https://www.salesforce.com/resources/articles/how-calculate-customer-churn-and-revenue-churn>
42. Silge, J. & Robinson, D. (2022). *Term Frequency and Inverse Document Frequency (tf-idf) Using Tidy Data Principles*. Retrieved 1 September 2022 from [https://cran.r-project.org/web/packages/tidymodels/vignettes/tf\\_idf.html](https://cran.r-project.org/web/packages/tidymodels/vignettes/tf_idf.html)
43. Slapničar, S. (2004). *Poročanje o družbeni odgovornosti [Reporting on corporate social responsibility]*. In S. Prašnikar (ed.), *Razvojnoraziskovalna dejavnost ter inovacije, konkurenčnost in družbena odgovornost podjetij [R&D and innovation, competitive advantage, and CSR]* (pp. 519–542). Ljubljana: Finance.
44. Statista. (2022). *ESG scores of the worlds' largest companies 2021*. Retrieved 7 September 2022 from <https://www.statista.com/statistics/1268534/comparison-esg-scores-largest-companies-provider-worldwide/>
45. Statista. (n.d.). *Customer service: Churn rate by industry U.S.* Retrieved 7 November 2022 from <https://www.statista.com/statistics/816735/customer-churn-rate-by-industry-us/>
46. Statistics Solutions. (n.d.). *Homoscedasticity*. Retrieved 17 September 2022 from <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/homoscedasticity/>
47. Statistični urad Republike Slovenije. (n.d.). *Pregled klasifikacije–Klasje*. Retrieved 7 November 2022 from <https://www.stat.si/Klasje/Klasje/Tabela/5531>
48. Stockopedia. (n.d.). *Net Debt to Assets Meaning*. Retrieved 14 September 2022 from <https://www.stockopedia.com/ratios/net-debt-to-assets-latest-487/>
49. Techslang. (2019, August 19). *What is Lemmatization?* Retrieved 1 September 2022 from <https://www.techslang.com/definition/what-is-lemmatization/>
50. Thomson Reuters. (n.d.). *Proxy Contest*. Retrieved 29 October 2022 from [https://uk.practicallaw.thomsonreuters.com/5-383-2214?transitionType=Default&contextData=\(sc.Default\)](https://uk.practicallaw.thomsonreuters.com/5-383-2214?transitionType=Default&contextData=(sc.Default))

51. Trgovinska zbornica Slovenije. (2022). *Trajnostno korporativno poročanje*. Retrieved 17 September 2022 from <https://www.tzslo.si/sl/vsebine-za-clane/splosna-zakonodaja/okolje/trajnostno-porocilo>
52. United Nations. (n.d.). *The 17 Goals*. Retrieved 7 September 2022 from <https://sdgs.un.org/goals>
53. University of Virginia Library. (n.d.). *Understanding Q-Q Plots*. Retrieved 17 September 2022 from <https://data.library.virginia.edu/understanding-q-q-plots/>
54. Wahyuni, P. D. (2020). Corporate governance and sustainability reporting practices: market dimension of financial performance. *International Journal of Business, Economics and Law*, 22(1), 9.
55. Wikipedia. (2022a). *TF-IDF*. Retrieved 7 September 2022 from <https://en.wikipedia.org/w/index.php?title=Tf%E2%80%93idf&oldid=1071253989>
56. Wikipedia. (2022b). *Tf-idf*. Retrieved 7 September 2022 from <https://en.wikipedia.org/w/index.php?title=Tf%E2%80%93idf&oldid=1071253989>
57. Wikipedia. (2022c). *Flesch–Kincaid readability tests*. Retrieved 7 September 2022 from [https://en.wikipedia.org/w/index.php?title=Flesch%E2%80%93Kincaid\\_readability\\_tests&oldid=1068065038](https://en.wikipedia.org/w/index.php?title=Flesch%E2%80%93Kincaid_readability_tests&oldid=1068065038)
58. Wikipedia. (2022d). *Flesch–Kincaid readability tests*. Retrieved 7 September 2022 from [https://en.wikipedia.org/w/index.php?title=Flesch%E2%80%93Kincaid\\_readability\\_tests&oldid=1101119398](https://en.wikipedia.org/w/index.php?title=Flesch%E2%80%93Kincaid_readability_tests&oldid=1101119398)
59. WordSense Dictionary. (n.d.). *Sustinere*. Retrieved 6 September 2022 from <https://www.wordsense.eu/sustinere>

## **APPENDICES**

## Appendix 1: Povzetek (Summary in Slovene language)

V zadnjih letih je opazno poraslo zanimanje za primerjavo finančne uspešnosti in trajnostnega poročanja podjetij po vsem svetu. Pomanjkanje tovrstnih raziskav v Sloveniji je eden od glavnih povodov za nastanek tega magistrskega dela.

Trajnost izvira iz latinske besede *sustinere*, kar pomeni "stati trdno" (*Sustinere*, n. d.). V poslovnem svetu se nanaša na zagotavljanje dolgoročnega uspeha ob hkratnem ohranjanju ali izboljševanju stanja okolja. Trajnostni razvoj pomeni zagotavljanje čistosti vode, zraka in zemlje ter zagotavljanje zadostnih zalog hrane za zadostitev osnovnih človekovih potreb. Podjetja poročajo o trajnosti v tako imenovanih trajnostnih poročilih, ki so posledično izvrsten odraz trajnosti oz. trajnostnega razvoja podjetja.

Trajnost je še vedno večkrat povezana z negativnim finančnim učinkom za podjetje. Z namenom potrditve oziroma ovržbe zgornje trditve, to magistrsko delo prikazuje povezave med kvaliteto trajnostnih poročil in finančne uspešnosti na vzorcu večjih slovenskih podjetij.

Kvaliteto trajnostnih poročil sem kvantificiral z dolžino, berljivostjo, odstotkom vsebovanih trajnostnih besed in sentimentom. Za kvantificiranje finančne uspešnosti sem uporabil nekaj najbolj standardnih metrik: ROA, ROE, EBIT MARGIN, EBITDA MARGIN, NET DEBT / EBITDA ...

Vse metrike in analize so bile pridobljene oz. pripravljene v programskem jeziku R oz. python. Uporabljeni so bili paketi Tesseract, pdftools, Stringr, NLP, tidyverse, tidytext, (python paket) lemmagen, quantenda, tidytext, Topic models. Za iskanje povezave med metrikami, ki predstavljajo vidik kvalitete trajnostnih poročil in metrikami finančne uspešnosti sem uporabil linearne in logistične regresijske modele.

Povezavo sem iskal z odgovori na naslednja raziskovalna vprašanja:

1. V kolikšni meri so učinkovita in dobro strukturirana trajnostna poročila podjetij povezana z dobro finančno uspešnostjo?
2. Kaj lahko iz vsebine trajnostnega poročila in njegovih lastnosti sklepamo o prihodnji finančni uspešnosti?
3. Kako je kakovost trajnostnega poročila odvisna od sektorja podjetja?

Rezultati raziskave pokažejo, da so obsežna in dobro strukturirana poročila povezana z boljšo finančno uspešnostjo, kar je v skladu z dosedanjo literaturo.

Dokazana je bila statistično značilna povezava med vsebino trajnostnih poročil in finančno uspešnostjo podjetja. Na vprašanje "Na kakšen način sta povezani vsebina trajnostnih poročil in finančna uspešnost?" pa je na podlagi naših rezultatov nemogoče odgovoriti zaradi med seboj izključujočih se ugotovitev. Na tem področju bodo torej potrebne še dodatne

raziskave v prihodnosti za točno opredelitev povezanosti, po možnosti z uporabo umetne inteligence.