



DALARNA
UNIVERSITY

Degree Thesis

Master's level (Second cycle)

Decoding Visual and Textual Elements in CSR Reports: A Systematic Analysis of Images and Text for Corporate Sustainability Insights

Author: Julian Weerasinghe and Nilupa Batawala
School: Dalarna University
Supervisor: Asif M. Huq and Vijay Pratap Paidi
Examiner: Mia Xiaoyun Zhao
Subject/main field of study: Microdata Analysis
Course code: MI4002
Credits: 15 Credits
Date of examination: 11th Jan 2024

At Dalarna University it is possible to publish the student thesis in full text in DiVA. The publishing is Open Access, which means the work will be freely accessible to read and download on the internet. This will significantly increase the dissemination and visibility of the student thesis.

Open Access is becoming the standard route for spreading scientific and academic information on the internet. Dalarna University recommends that both researchers as well as students publish their work Open Access.

I give my/we give our consent for full text publishing (freely accessible on the internet, Open Access):

Yes

No

Dalarna University – SE-791 88 Falun – Phone +4623-77 80 00

Abstract:

This thesis examines the interplay of visual and textual discourse in Corporate Social Responsibility (CSR) reports, offering a systematic framework to analyse a dataset comprising around 66,925 images from 675 CSR reports. By analysing image attributes, colours, and objects in conjunction with textual sentiment and topics, we investigate the similarities, contrast and trends across various sectors and regions, and the impact of company characteristics. The mixed-methods approach, incorporating both qualitative image analysis and quantitative text evaluation, reveals patterns in how CSR initiatives are visually and textually communicated. Image and text extraction were accomplished using PyMuPDF and Tesseract libraries, harnessing the OCR capabilities. The identification of living objects was performed using OpenCV, while image classification was executed with the OpenAI-CLIP model, yielding high accuracy in extracting the visual content of the images. The developed framework achieved accuracy rate of 81% on living object identification using OpenCV model and 76% accuracy in object classification using OpenAI-CLIP model. The study's results indicate that the distinct patterns in how CSR is depicted, varying by sector, geographic location, and company size. These patterns offer key insights for developing more targeted and effective strategies for engaging with stakeholders.

Keywords:

Corporate Social Responsibility (CSR), Data Mining, Image Analysis, Textual Analysis, Object Extraction, Image Classification,

Table of Contents

1 Introduction	1
2 Literature Review	4
2.1 Image object attribute classification in CSR reports	4
2.2 Sentiment analysis and topic modelling in CSR	6
3 Methodology	7
3.1 Dataset Refinement	10
3.2 Image and Text Extraction	15
3.3 Image Attribute Extraction and Object Classification	19
3.3.1 Colour Extraction	20
3.3.2 Living Object Extraction and Object Classification	22
3.4 Text Preprocessing	27
3.5 Sentiment Analysis	27
3.6 Topic Modelling	29
3.7 Data Indexing and Integration	30
3.8 Exploratory Data Analysis (EDA)	31
3.9 Manual Sample Verification	35
4 Results	36
4.1 Sector based findings	39
4.2 Company Size based findings	46
4.3 Company Region based findings	47
5 Discussion	50
5.1 Limitations of the Study	51
5.2 Recommendations for Future Research	52
6 Conclusion	53
References	55
Appendix 1 – Results of Manual Verification on Object Identification and Object Classification	60
Appendix 2 – Dominant Topic and Dominant Colour based findings	63
Appendix 3 – Thesis work plan and individual contribution	66

List of Figures

Figure 1: Sequential steps of the methodology.....	8
Figure 2: Density of green, red, and blue colour percentage of images.....	21
Figure 3: Density of classified object probabilities	25
Figure 4: Density graphs of sentiment scores for positive and negative categories in sector groups	29
Figure 5: Density of sentiment scores for dominant objects	29
Figure 6: Analysis Strategy	33
Figure 7: Dominant classified object type percentage in sector groups	40
Figure 8: Time series analysis for dominant object percentage in sector group....	41
Figure 9: Dominant objects distribution among dominant topics for sector groups.....	42
Figure 10: Dominant colour distribution in sector Groups	43
Figure 11: Correlation between sentiment and dominant colour in sector groups.....	44
Figure 12: Correlation between classified object probability and positive sentiment in sector groups	44
Figure 13: Correlation of classified object probability and negative sentiment in sector groups	45
Figure 14: Average sentiment score distribution with presence of living object in sector groups	45
Figure 15: Dominant classified object type percentage in company size	46
Figure 16: Dominant objects and topic distribution in company size	47
Figure 17: Dominant classified object type percentage in regions	48
Figure 18: Dominant objects and topic distribution in region	49
Figure 19: Distribution of average sentiment score with presence of living objects in regions	49

List of Tables

Table 1: CSR Report/Image Statistics	12
Table 2: CSR reports/images statistics per year and sector	13
Table 3: Dataset fields	14
Table 4: HSV colour ranges	21
Table 5: Classified Object Category	25
Table 6: Extracted and defined attributes for the analysis	30
Table 7: Sector Groups	34
Table 8: Configuration of the personal computer program executed	35
Table 9: Summary of Manual Verification results	37
Table 10: Comparison of max probability derived from CLIP model for Objects categories with presence of Living objects	38

Acknowledgement

We extend our heartfelt thanks to our supervisors, Asif M. Huq and Vijay Pratap Paidi, for their continual guidance and feedback throughout our thesis. Without their support, completing this work on time would not have been possible. Overcoming numerous challenges and hurdles was made easier by our supervisors' encouragement, which helped us maintain focus and achieve our goals.

Finally, we offer special thanks to our families and friends. Their unwavering presence was a significant strength to us during this journey.

Acronyms

CSR (Corporate Social Responsibility)

CSRD (Corporate Sustainability Reporting Directive)

ESG (Environmental, Social Governance)

EU (Europe Union)

VADER (Valence Aware Dictionary and sEntiment Reasoner)

GRI (Global Reporting Initiative)

OCR (Optical Character Recognition)

HSV (Hue, Saturation, Value)

CLIP (Contrastive Language-Image Pre-training)

CSRW (Corporate Social Responsibility Words)

MNE (Multinational Enterprises)

SME (Small & Medium Enterprises)

Agri (Agriculture)

Tech (Technology)

Telco (Telecommunication)

Auto (Automobile)

1 Introduction

Corporate Social Responsibility (CSR) has become a critical element of corporate governance, reflecting a firm's commitment to ethical behaviour, environmental stewardship, and social welfare. CSR reporting, traditionally a voluntary practice, enables businesses to showcase their sustainability efforts and ethical conduct beyond mere financial metrics. The introduction of the Corporate Sustainability Reporting Directive (CSRD) on January 5, 2023, signifies a significant shift in this area (European Commission, 2023).

Under this new directive, Europe Union (EU) businesses, as well as qualifying EU subsidiaries of non-EU companies, are mandated to report on the environmental and social impacts of their activities, as well as the business implications of their environmental, social, and governance (ESG) efforts (IBM, 2023). This mandate elevates the importance of CSR communication which is vital for companies to showcase their commitment towards social and environmental issues, address stakeholder expectations, and build trust and credibility with various stakeholders.

CSR communication helps organizations respond to changing stakeholder expectations and societal standards, while CSR reporting provides transparency and accountability by sharing information about environmental actions and social responsibility. Analysing and evaluating CSR communication and reporting can provide insights into the effectiveness of an organization's CSR activities, identify areas for improvement, and enhance the overall impact of their sustainability efforts. Moreover, CSR reporting serves as a strategic tool for companies (Disability as a Blind Spot in Sustainability Frameworks and Standards, 2019).

It is important to consider both text and images when analysing CSR reports to get a comprehensive understanding of the organization's CSR initiatives and their impact, since text provides information about CSR initiatives, strategies and outcomes and images can complement this by visually demonstrating the impact of these initiatives, showing real world applications, or highlighting the people and

communities affected. Together they offer a more holistic view of company's CSR reports. In a discussion about effective corporate social responsibility (CSR) communication, the Vodafone report is cited as a positive example. It demonstrates how the use of images and text can be aligned to convey a unified message about the company's commitment to enhancing society through communication technologies. Similarly, Deutsche Telekom's report is noted for its coherent storytelling, effectively combining images and text. Titled "Human Resources and Sustainability Report," it showcases the significant role of employees in the company's non-financial achievements. (Breitbarth et al., 2010)

In previous research on CSR reports, most of the focus was on studying the textual content. Not many studies looked at the images, even though images can make a report more interesting and create effective communication. In the realm of image analysis, traditional approaches have predominantly involved manual coding and qualitative methodologies, which are impractical for extensive datasets. When incorporated into content analysis research, images have typically been subjected to manual coding processes or interpreted subjectively, based on personal perspectives (Hellmueller & Zhang, 2019; Serafini & Reid, 2019).

CSR reports often use a variety of images in different formats and styles, making systematic comparison and analysis difficult. To truly understand these images, they need to be considered alongside the text, adding complexity since extracting both images and adjacent text together and explore their interrelationship is a challenging task. The volume and diversity of images which consists of multiple layers along with embedded text in CSR reports also pose challenges, as manual analysis is time-consuming. Hence, there's a growing need to develop a framework to help stakeholders analyse these images effectively and understand their impact on perceptions of a company's CSR activities.

The main goal of this research is to analyse and quantify how the attributes of images in CSR reports are related to the topics and sentiment identified in the text surrounding these images. This analysis will also explore how the relationship of

image and text attributes differ when considering variables such as company size, geographic location, and industry sector. The thesis is anchored on the crucial question: **How are the attributes of images in CSR reports associated with the sentiment and thematic content of adjacent text, and in what ways do these associations vary across companies of diverse sizes, locations, and sector?** This investigation aims to shed light on the complexities of visual communication within CSR reports, examining how it complements or contrasts with the textual narrative presented by companies from various sectors, regions, and locations. By exploring the above-mentioned research question, we expect to achieve below three objectives.

- 1. Develop and implement a systematic framework for extracting image attributes and associated text, employing data mining techniques to generate valuable insights for corporate sustainability analysis.**
- 2. Investigate the utilization of image attributes in CSR reports and explore their trends with company characteristics, including size, location, and sector.**
- 3. Analyse sentiments and topics within CSR reports, exploring their connections with the attributes of images featured in these reports.**

The study seeks to establish a systematic framework for analysing the qualitative content of images in CSR reports and correlate with the topics and sentiment of the adjacent text. In the wake of new CSRD regulations, this research endeavours to contribute to the evolving discourse on corporate transparency, inform policy development, and assist companies in effectively communicating their CSR endeavours to stakeholders, ultimately driving forward the agenda of sustainable and ethical business practices.

The use of a systematic framework will allow for consistent and objective analysis of images, enabling researchers to gain deeper insights into the visual representation of CSR efforts. By analysing images in CSR reports, researchers can gain a better understanding of how companies communicate their CSR initiatives visually and

the potential impact these images may have on stakeholder perceptions and attitudes. Additionally, analysing images in CSR reports can help identify trends and patterns in visual storytelling techniques used by companies to convey their CSR activities.

The structure of the thesis is organized as follows: Section 2 offers a literature review, followed by Section 3, which outlines the research methodology. Section 4 details the results of the data analysis. Discussions of these results and suggestions for future research are presented in Section 5. Finally, Section 6 concludes the thesis by addressing the research questions.

2 Literature Review

The objective of the literature review is to examine past research studies to gain insights into the tools and methodologies used to analyse text and images with respect to image attribute extraction, topic modelling and sentiment analysis.

2.1 Image object attribute classification in CSR reports

The analysis of image objects in CSR reports plays a crucial role in understanding the visual representation and communication strategies employed by companies. Past studies in CSR report analysis have highlighted the limitations of solely relying on textual content analysis. These limitations include the inability to capture the visual impact and emotional appeal of images, as well as the difficulty in quantifying subjective elements such as aesthetics and perceptions. The aim of this thesis is to overcome these limitations by incorporating both textual content analysis and image object analysis in the evaluation of CSR reports. From the legitimacy perspective, the use of images and photos in CSR communication plays a fundamental role because it can be implemented to strategically improve stakeholders' perceptions of the organization (Invernizzi et al., 2021) Additionally, analysing the image objects allows for a deeper understanding of how companies shape their corporate image and visually depict their social and environmental practices. This comprehensive approach will provide a more holistic understanding of the communication

strategies used in CSR reports and contribute to the broader field of CSR research. By integrating textual content analysis and image object analysis, this thesis aims to provide a more comprehensive understanding of the visual communication strategies employed by companies in their CSR reports.

Image in CSR reports may contain object such as human, animal, vehicle, building or nature elements. Several studies have focused on the classification of objects in CSR reports using various techniques. One study by Luo et al. (2023) proposed a subject detection strategy combined with image categorization to classify objects in CSR reports. They introduced a different subject/background segmentation method and extracted visual features based on the class of the image. By employing this approach, they were able to effectively categorize different objects in CSR reports and extract relevant visual features for analysis.

Another study conducted by Dupire and M'Zali focused on the relationship between visual impression management and CSR performance in CSR reports (Anantharaman, Huang & Zhao, 2020). The researchers randomly selected 200 CSR reports and classified the 10,596 images they contained into aesthetic and non-aesthetic categories. They refined their documentation of image usage in CSR reports by introducing a classification method and evaluating its effectiveness. The aim of these studies was to explore the use of image object attribute classification in CSR reports. Another study by Liu applied CSR (Convolutional Sparse Representation) to image fusion and compared it to SR-based methods (Wang et al., 2021). The results showed that the CSR-based fusion method had better evaluation results and visual effects.

Overall, image object attribute classification is an effective methodology for analysing and understanding the visual content of these CSR reports. By classifying the objects in CSR reports, researchers and analysts can gain insights into the types of images being used, the visual impression being created, and the potential impact on CSR performance and reputation management. In summary, image object

attribute classification in CSR reports enables researchers and analysts to gain valuable insights into the visual content of these reports.

2.2 Sentiment analysis and topic modelling in CSR

Important aspect of CSR reporting is the use of sentiment analysis and topic modelling techniques. Sentiment analysis, also known as opinion extraction or review mining, refers to the process of identifying and extracting subjective information from text using natural language processing and artificial intelligence (Duraivel & Lavanya, 2022).

Sentiment analysis in CSR reports involves analysing the sentiment or tone of the language used in these reports to determine whether it is positive, negative, or neutral. Sentiment analysis can be used to assess public perception and response to a company's CSR initiatives, providing insights into stakeholders' feelings and attitudes towards the organization and its social and environmental practices. Furthermore, sentiment analysis can help identify any discrepancies or inconsistencies between a company's CSR performance and the tone of its CSR reporting. Several studies have applied sentiment analysis and topic modelling techniques to analyse CSR reports. For example, Mućko et al. conducted a study to evaluate the sentiment of CSR disclosures in annual reports of EU companies (Mućko, 2021). They used sentiment analysis to determine the positive or negative sentiment expressed in the reports and found a significant relationship between CSR disclosure sentiment and financial performance. The researchers used the Lexicoder Sentiment Dictionary and the Corporate Social Responsibility content analytic dictionary to identify related mission sentiment and topics related to societal relevance in the CSR reports (Cortés-Sánchez, 2022).

Sentiment analysis is a valuable tool in analysing and understanding the sentiment expressed in corporate social responsibility reports. By examining the emotional tone and attitudes conveyed in these reports, sentiment analysis can provide insights

into how stakeholders perceive a company's CSR initiatives (Sung, Tao & Slevitch, 2022).

One widely used sentiment analysis tool in this context is VADER (Moses, 2021). It stands for Valence Aware Dictionary and sEntiment Reasoner (Muttineni & Deng, 2021). VADER is a lexicon and rule-based sentiment analysis tool specifically designed for social media texts, making it well-suited for analysing sentiment in CSR reports that may include language and expressions commonly found on social media platforms. VADER is unique in its approach to sentiment analysis as it takes into account not only the sentiment of individual words, but also word-order relationships and specific lexical features such as abbreviations and slang.

3 Methodology

Our methodology was carefully devised to comprehensively address the three stated objectives outlined in the introduction section of this thesis report. Previous studies highlight the strategic role of images as a means of shaping corporate impressions and managing perceptions (Anantharaman et al., 2020). Given the multifaceted nature of the research question, a combination of qualitative image analysis and quantitative measures of CSR disclosure quality is deemed most suitable. The study involves an in-depth analysis of images and associated text extracted from non-financial reports in the global context. This includes the identification of different layers within each image, extracting text within and around the images, extracting image attributes and classifying images based on presence of different object categories. This research adopts a mixed-methods approach, leveraging both qualitative and quantitative methodologies. Given the multifaceted nature of the research question, a combination of qualitative image analysis and quantitative measures of CSR disclosure quality is deemed most suitable. The study involves an in-depth analysis of images and associated text extracted from non-financial reports in the global context. This includes the identification of different layers within each image, extracting text within and around the images, extracting image attributes and classifying images based on presence of different object categories. To complement

the qualitative findings, quantitative measures are employed. This involves classifying images into distinct categories and extracting attributes such as sentiment analysis and topic modelling from the associated text. These measures aim to provide a quantitative basis for understanding the thematic dimensions of CSR reports regarding correlation of the images and associated text. The provided flowchart illustrates the sequential steps undertaken in our research.

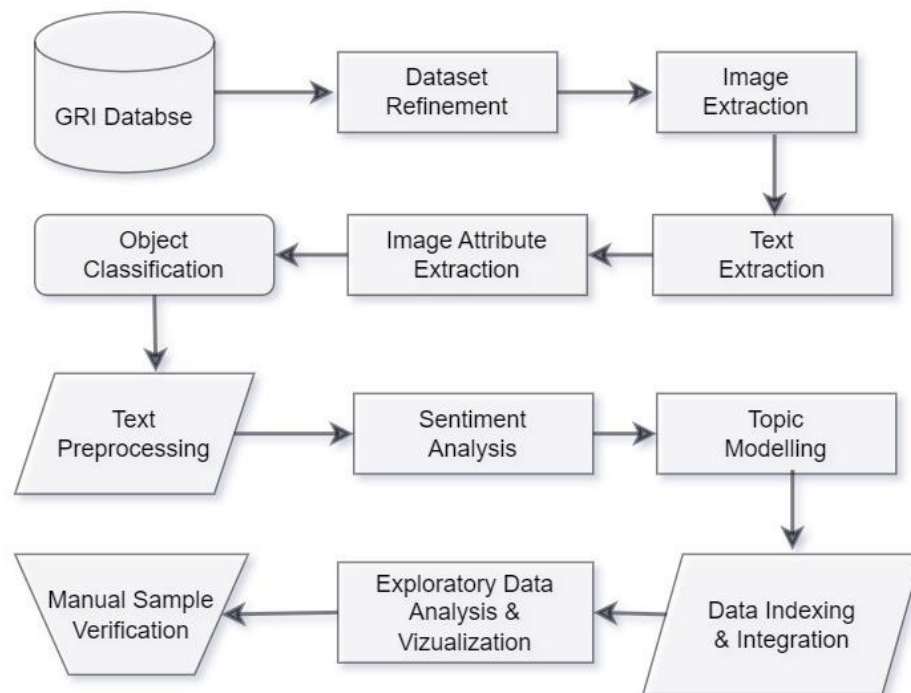


Figure 1: Sequential steps of the methodology

In pursuit of our research objectives, we crafted a methodology encompassing the following sequential steps demonstrated in Figure 1.

1. **Dataset Refinement:** We initiated the process by filtering the dataset according to a defined criteria to organize an interpretable and convenient analysis strategy. This involved a targeted selection approach, where emphasis was placed on specific sectors aligned with our research focus. The

criteria included sectorial relevance, adherence to the English language, and compliance with GRI reporting standards.

2. **Image Extraction:** Employing Python libraries, we extracted images from selected CSR reports. This step was crucial for capturing diverse layers associated with distinct image compositions. The resulting images were systematically saved in universally compatible formats of JPG and PNG.
3. **Text Extraction:** Utilizing state-of-the-art Optical Character Recognition (OCR) libraries in Python, we extracted textual layers embedded within and surrounding the images. This comprehensive approach ensured the capture of textual content crucial for subsequent analyses.
4. **Image Attribute Extraction and Object Classification:** We proceeded objects within the images. Leveraging sophisticated techniques, this step aimed to unravel the visual elements and categorize objects, laying the groundwork for further exploration, which is aligned with the 1st objective of our thesis.
5. **Text Preprocessing:** Before delving into sentiment analysis and topic modelling, a crucial precursor involves rigorous text preprocessing. This multistep process is designed to refine the raw text data, ensuring optimal input for subsequent analyses.
6. **Sentiment Analysis:** The sentiment of the text accompanying each image was calculated with precision. This involved an in-depth assessment of the textual content to ascertain its polarity, providing valuable insights into the emotional tone associated with the corporate disclosures.
7. **Topic Modelling:** A robust topic modelling process was employed to delve into the thematic content of the text associated with each image. This step facilitated the identification and categorization of prevalent topics within the CSR reports.
8. **Data Indexing and Integration:** Each image was uniquely indexed, and the extracted attributes, including image features, sentiment scores, and topics,

were systematically merged with the corresponding index. This indexing allowed for a seamless integration of diverse data elements.

9. **Exploratory Data Analysis (EDA) and Result Visualization:** We conducted an in-depth Exploratory Data Analysis, aligning with a predefined analysis strategy crafted to address our overarching research question. This phase focused on uncovering patterns, trends, and correlations within the dataset. The results of the exploratory and comparative analyses were visualized, providing a clear and insightful representation of the relationships between image attributes, sentiments, and topics. Visualization served as a powerful tool for communicating the findings derived from our methodology.
10. **Manual Verification of Automated Classification:** We executed a manual verification process for the extraction and classification of image attributes to assess the performance of the machine learning algorithm. This verification involved a comparison between the results obtained from the machine learning steps of living object detection and object classification with the outcomes determined by human observation. A random sample of 200 images was selected for this comparative analysis.

This comprehensive methodology was purposefully designed to align with the complexities of our research goals, ensuring a systematic and thorough exploration of the relationships between visual and textual elements in CSR reports. Each of the steps are detailed further in the subsequent subsections found within the methodology section.

3.1 Dataset Refinement

This study relied on the GRI database and reports that were downloaded from Lindgren et al (2021). The repository offered an extensive compilation of non-financial reports across diverse sectors and countries, available in PDF format. The GRI database provided essential company attributes, including company size, sector, country, and language of the report. Company size was used to differentiate

between large enterprises, multinational corporations, and medium/small enterprises.

While the original dataset consists of 12 sectors, the dataset underwent a targeted selection process, specifically concentrating on sectors related to agriculture, food, automotive, and technology-related sectors. This deliberate focus on these sectors enhances the precision and interpretability of our analysis within the defined scope of this research. This considered selection of sectors within our dataset was a deliberate and strategic process geared towards aligning with the overarching research question. By choosing these specific sectors, we aim to capture a diverse spectrum of industries that are inherently contrasting in their nature. For instance, agriculture and technology are industries with distinctive characteristics, often diverging in their visual narratives and corporate communication strategies (Buallay, 2021). Agriculture, with its emphasis on sustainability, environmental impact, and community engagement, may manifest visual attributes that differ significantly from the technology sector, which may highlight innovation, technological advancements, and global connectivity (Jackson et al., 2019). Drawing upon existing literature and studies within these sectors, we anticipate encountering observable differences in the types of images used, reflective of each sector's unique priorities, challenges, and communication objectives. By exploring these contrasting sectors, we expected to enhance our ability to unravel variations in image attributes and their associations with sentiment and thematic content in CSR reports.

Criteria for Sample Selection:

The selection of CSR reports was guided by specific criteria aimed at aligning with the research objectives. The following criteria were applied which resulted 958 reports out of 19,086 reports

1. Sectoral Focus: Reports within the agriculture, food, automotive and three technology related sectors were prioritized to ensure inclusion of diversified

sectors for the analysis. The technology sector encompasses a diverse range of industries, including Equipment, Technology Hardware, and Telecommunication sub-sectors.

2. Language Filter: Only reports in the English language were considered to maintain the linguistic consistency in the dataset.
3. GRI Version: Reports adhering to GRI versions 3, 3.1, and 4 were selectively included, aligning with widely recognized reporting standards and facilitating a cohesive analysis. GRI certified training partner program introduced in 2008 and CSR reports that compliant with GRI 3 to 4 versions in the dataset are standardized by this certification (GRI, 2022). Such standardisation in selected sample is important to improve the quality of the results.

Table 1: CSR Report/Image Statistics

Population	19086	Total reports - images extracted	946
		Reports failed to extract images	12
Sample	Report Count		
Agriculture	44	Report count failed to extract image attribute	271
Food & Beverage	230	Final report counts for analysis	675
Automotive	111	Total number images extracted	67636
Equipment	116	Total na values removed	711
Technology Hardware	230	Final image count for analysis	66925
Telecommunication	227		
Total reports	958		

The Table 1 highlights that for this study, six sectors were chosen: Agriculture, Food & Beverage, Automotive, Equipment, Technology & Hardware, and Telecommunications out of 38 sectors. The reason we selected these sectors is to reduce the variability of the sample and enhancing the interpretability of the results. The selected sectors represent contrasting domains and support us to conduct a comparative analysis between these sectors. These sectors collectively encompassed 958 CSR reports from a total of 19,086. From these reports, images were successfully extracted from 946, while extraction failed for 12 due to various issues such as unresponsive image formats and corrupted images. Subsequently, image attributes were extracted from the obtained images for each report. This process was successful for images associated with 675 reports, but it failed for 271 reports due to image corruption.

Table 2: CSR reports/images statistics per year and sector

Publication Year	Agri & Food image count	Agri & Food report count	Tech & Telco image count	Tech & Telco report count	Auto & Equipment image count	Auto & Equipment report count	Total Images Count Per Year	Total CSR Report Count Per Year
2006	104	1	39	1	0	0	143	2
2007	111	1	420	2	0	0	531	3
2008	233	3	545	4	20	1	798	8
2009	668	6	428	7	205	4	1301	17
2010	546	6	883	11	822	6	2251	23
2011	1141	10	1218	21	262	5	2621	36
2012	1350	10	3176	16	994	13	5520	39
2013	1708	21	4268	47	1495	15	7471	83
2014	4817	35	3623	50	1653	26	10093	111

2015	3643	37	4252	47	3481	28	11376	112
2016	4340	44	5308	63	2864	24	12512	131
2017	4346	37	4920	45	2773	25	12039	107
2018	169	1	33	1	67.0	1	269	3
Total	23176	212	29113	315	14569	148	66925	675

The Table 2 shows the CSR report file count and image distribution per year from 2006 to 2018. It demonstrates the year-[wise](#) sectorial contribution of CSR reports and images to our study.

The data fields utilized in this analysis are presented in Table 3, illustrating whether each field originated from the original data source or was extracted through our analysis methodology steps. A comprehensive discussion regarding the values represented in the extracted data fields and their definitions will be provided in the subsequent subsections of the methodology.

Table 3: Dataset fields

Data Field	Source	Data Type	Description
File_Name	GRI dataset	String	CSR Report File Name
Image_Name	Extracted from Reports	String	Extracted Image name (unique index)
Text_Before	Extracted from Reports	String	Extracted text before the image
Text_After	Extracted from Reports	String	Extracted text after the image
Text_Inside_Image	Extracted from Reports	String	Extracted text inside image
green_perc	Extracted from Reports	Numeric	Green percentage extracted from the image. Explained in section 3.3.1
blue_perc	Extracted from Reports	Numeric	Blue percentage extracted from the image. Explained in section 3.3.1
red_perc	Extracted from Reports	Numeric	Red percentage extracted from the image. Explained in section 3.3.1
live_obj	Extracted from Reports	Boolean	If a living object identified value set to 'TRUE', if not set to 'FALSE'. Explained in Section 3.3.2

Human	Extracted from Reports	Numeric	Probability of the image attribute identified is human. Explained in Section 3.3.2
Animal	Extracted from Reports	Numeric	Probability of the image attribute identified is animal. Explained in Section 3.3.2
Nature	Extracted from Reports	Numeric	Probability of the image attribute identified is nature. Explained in Section 3.3.2
Vehicle_Or_Machine	Extracted from Reports	Numeric	Probability of the image attribute identified is vehicle or machine. Explained in Section 3.3.2
Other	Extracted from Reports	Numeric	All other objects do not classify to either human, animal, vehicle or machine. Explained in Section 3.3.2
Sentiment_Score_text_in_around_image	Extracted from Reports	Numeric	Generated sentiment score of the around the image and inside image. Explained in Section 3.3.5
Size	GRI Dataset	String	Organization Size
Sector	GRI Dataset	String	Organization Sector
Region	GRI Dataset	String	Organization Region
Publication_Year	GRI Dataset	Numeric	Publication Year of the report
CSRW - Employee	Extracted from Reports	Numeric	Topic count for Employee category. Explained in Section 3.3.6
CSRW - Environment	Extracted from Reports	Numeric	Topic count for Environment category. Explained in Section 3.3.6
CSRW - Human Rights	Extracted from Reports	Numeric	Topic count for Environment category. Explained in Section 3.3.6
CSRW - Social and Community	Extracted from Reports	Numeric	Topic count for Environment category. Explained in Section 3.3.6
group_sector	GRI Dataset	Numeric	Group Sectors 'Agri & Food', 'Auto & Equipment', 'Tech & Telco'

3.2 Image and Text Extraction

The process of data extraction was a crucial phase, as the attributes obtained from both images and text serve as the fundamental building blocks for our research. Our approach was to extract all images initially from the selected reports based on predefined criteria. To establish a systematic methodology, we chose to employ Python program for extracting images and associated text. The aim of the image and text extraction step is twofold: firstly, to store each image independently in distinct folders, and secondly, to compile the associated text in a dataset. Aligning with the research objective, our goal was to extract 10 lines of text both preceding and following the report, in addition to capturing text within the images themselves. This

approach ensures a comprehensive collection of contextual information surrounding each image, contributing to the richness of the dataset for subsequent analysis.

Traditional Python libraries for image and text extraction such as Python Imaging Library (PIL), NumPy fell short of delivering satisfactory results, primarily due to the following factors.

1. Diverse Image Formats:
 - The presence of images in various formats posed a challenge for traditional extraction libraries.
2. Varied Layered Structures:
 - Images exhibited diverse layered structures, making it challenging to extract content consistently.
3. Text-Embedded Images:
 - Images were embedded with text, requiring an advanced approach to extract both visual and textual components.
4. Abundance of Image Collections:
 - Some images comprised large collections, necessitating an effective strategy to handle extensive datasets.

The implemented method for extracting images and texts from PDF reports crucially leverages the PyMuPDF and Tesseract libraries in Python. PyMuPDF facilitates the parsing of PDF documents, enabling access to individual pages and images within each page. However, due to the diverse nature of image formats, layered structures, and the presence of embedded text in the reports, traditional Python libraries for image extraction proved inadequate for our research needs. This led to the adoption of Tesseract, an open-source Optical Character Recognition (OCR) tool, to overcome these challenges. Tesseract stands out for its OCR capabilities, making it particularly adept at extracting text from images. The integration of Tesseract with PyMuPDF allowed us to convert image data to a format compatible with Tesseract, facilitating the extraction of text embedded within images. Moreover, Tesseract's

ability to recognize text even within images with complex layered structures addressed the challenge posed by varying image compositions (Akinbade et al, 2020.).

In certain instances, challenges arose where a substantial number of images were consolidated into a single image, resulting in extended extraction times. Additionally, these small images were often limited to a few pixels in size. To address this, a strategic control measure was introduced during the PDF page extraction process. To avoid extraction of large number of small noisy images that contained in CSR reports it was required to introduce a validation before extract images from each page in CSR reports. Therefore, if a page contained more than 20 images, it was deliberately omitted from the extraction procedure. This control mechanism played a pivotal role in optimizing the extraction process by mitigating the potential delays associated with processing an extensive array of small images within a single page. By excluding the larger number of small images, this approach not only enhanced the efficiency of the extraction process but also contributed to the overall success of the research by omitting noisy images from the dataset. The exclusion of pages with an excessive number of images ensures a more streamlined and effective extraction, laying a solid foundation for later phases of analysis. This exclusion approach was vital for making the extraction process efficient, and establishing a comprehensive understanding of the relationship between image attributes, sentiments and topics derived from the textual content in subsequent stages of the research

Below are the key facts we consider when we extract the image and text extraction step of our methodology related to our research objectives.

1. Contextual Richness: By extracting both image attributes and surrounding textual content, the research aims to capture a more comprehensive view of the information presented in CSR reports. Images and text often complement each other, and understanding their relationship can provide richer contextual information.

2. **Holistic Analysis:** CSR reports are multimodal, containing both visual and textual elements. Analysing only one aspect might lead to an incomplete understanding. The combined analysis allows for a more holistic examination of the content, considering both the visual representation and the narrative provided in the text.
3. **Exploratory Analysis and Validation:** Analysing image attributes alongside sentiments and topics in the text enables the exploration of trends. It helps validate whether the visual elements align with the textual message or if there are discrepancies. This cross-validation enhances the reliability of the findings.
4. **Interdisciplinary Insights:** The integration of image attributes and text bridges the gap between visual analysis and natural language processing. This interdisciplinary approach allows researchers to draw insights from both computer vision and linguistic perspectives, contributing to a precise understanding of the content.
5. **Stakeholder Perception:** Stakeholders, including investors and the public, interpret CSR reports in different ways. The combined analysis of images and text can shed light on how visual elements influence the perception of corporate social responsibility initiatives. This is crucial for companies aiming to convey a positive image.
6. **Research Objectives Alignment:** The approach aligns with the research objectives outlined in the methodology. If the goal is to understand the impact of images on stakeholders' interpretation and the perceived authenticity of non-financial corporate disclosure, a thorough analysis of both image attributes and textual sentiments/topics is essential.
7. **Quality and Accuracy:** Addressing challenges related to image formats, layered structures, and embedded text requires a systematic approach to maintain data quality and accuracy. This ensures that the extracted information is reliable for subsequent phases of the research.

For each image, the extraction process involves determining the file extension based on image type, extracting text content before and after the image, and utilizing Tesseract to extract text within the image itself. The extracted images are then saved separately in designated folders in both 'jpg' and 'png' formats, while the associated text is stored in a dataset. By utilizing Tesseract's robust OCR capabilities, we successfully navigated challenges posed by different image formats, layered structures, and text-embedded images. Text and image extraction using Tesseract-OCR not only ensures the comprehensive extraction of visual and textual elements but also contributes to the formation of a rich dataset, laying the groundwork for subsequent in-depth analysis in our research.

3.3 Image Attribute Extraction and Object Classification

In the pursuit of understanding the impact of images within CSR reports on stakeholders' interpretation of non-financial corporate disclosure, the extraction of image features emerges as a pivotal aspect of this research. The significance lies in the ability to analyse and interpret the visual output of these images, shedding light on the expressions encapsulated within them. The extraction of image attributes is therefore integral to unravelling the role and influence of visual elements, providing valuable insights into how stakeholders perceive and engage with non-financial corporate disclosure. This process serves as a foundational step in comprehensively examining the interplay between visual content and stakeholder interpretation in the context of CSR reporting.

Upon inspecting the image attributes, a discernment emerged that three primary components could be extracted from the images: Colours, Objects, and Texts. Notably, Texts had already been extracted in the preceding step. Consequently, our focus in this phase shifted towards extracting the colour presentations and identifying the types of objects within the images. This strategic focus is driven by the recognition that the visual impact of images extends beyond textual content alone. The colour palette and the types of objects depicted in these images convey

additional layers of meaning and contribute significantly to stakeholders' interpretation and the overall delivery of non-financial corporate disclosure.

3.3.1 Colour Extraction

In the process of extracting image attributes, we specifically targeted colour presentations, focusing on the percentages of green, red, and blue hues within each image. We selected these three colors as an initial step of implementing color extraction in our framework since extracting and analyzing all the colors is not practical due to time constraints. This task was accomplished through a Python program using OpenCV library. The implemented methodology involves converting the image to the HSV (Hue, Saturation, Value) colour space, allowing for more effective colour segmentation. Subsequently, a predefined range for each colour (green, red, and blue) was established to create masks that isolate pixels falling within the specified colour range. The masks facilitated the calculation of the percentage of pixels representing each colour in relation to the overall image.

In the process of extracting colour information from images, we established standardized lower and upper ranges in the HSV colour space for each colour—red, green, and blue depicted in Table 4. These ranges serve as criteria to determine whether each pixel in the image possesses Hue, Saturation, and Value (brightness) values falling within the specified ranges for each colour. For instance, in the case of green colour extraction, the program identified the percentage of green pixels by comparing the number of non-zero pixels in the green mask to the total number of pixels in the image.

Table 4 below provides defined standard ranges for each colour in the HSV colour space (Rebecca Stone, 2018).

Table 4: HSV colour ranges

Colour	Lower Range (HSV)	Upper Range (HSV)
Green	Hue: 36, Sat: 25, Val: 25	Hue: 86, Sat: 255, Val: 255
Red	Hue: 0, Sat: 100, Val: 100	Hue: 10, Sat: 255, Val: 255
Blue	Hue: 100, Sat: 50, Val: 50	Hue: 130, Sat: 255, Val: 255

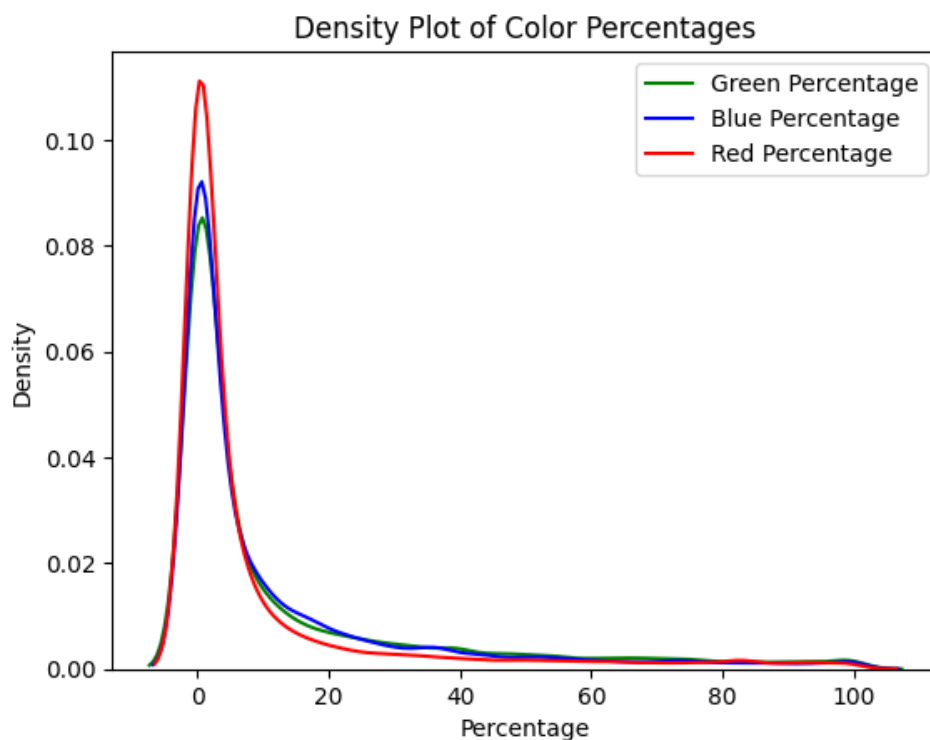
**Figure 2: Density of green, red, and blue colour percentage of images**

Figure 2, the density plot shows the distribution of colour percentages for green, blue, and red colours. in our dataset. The peaks are quite close to each other, indicating that the distribution of percentages for each colour is somewhat similar. The sharp peak and quick drop-off imply that most values are clustered around a lower percentage range, with relatively few instances of higher percentages for all three colours. In our analytical discussions, one of approaches was to identify the

predominant colour by ascertaining the highest recorded percentage of colour for each image. We utilize the term "dominant colour " to signify this identified colour.

3.3.2 Living Object Extraction and Object Classification

Object extraction posed a unique challenge in this research, considering the multitude of the machine learning approaches available for this task. Crucially, the determination of object types extracted from images was a key aspect, particularly in establishing relationship between these objects and the sentiments and topics identified in the accompanying text. Transitioning from a business-centric to a society-centered perspective has important implications for future CSR research, particularly in understanding cause-and-effect relationships and linking CSR strategies to actual social and environmental improvements (Wickert, 2021). Adopting a society-centric approach in CSR images involves incorporating human, animal, and nature objects to convey a company's dedication to societal well-being, nature, and human rights. This shift prompted our investigation into the presence of liveness in each image. For this purpose, the OpenCV library was employed, utilizing a dedicated function designed to detect the existence of living objects within an image. The function leveraged a pre-trained living Object detection model, specifically the 'haarcascade_frontalface_default.xml' file provided by OpenCV (Ashwini et al., 2020). By applying this model, the function could identify the presence of Living objects (data field name: live_obj), human or animal face within an image (Hasan & Sallow, 2021). The underlying objective was to ascertain whether alive objects, were present in the images.

Subsequently, we wanted to categorize images into different object types using pre-defined object labels. The aim was to discern the dominant object categories to which images could be assigned based on their visual representations. The solution was found in leveraging OpenAI's pre-trained model, the CLIP (Contrastive Language-Image Pre-training) model with the architecture "openai/clip-vit-large-patch14." This refined approach involved utilizing the CLIP model in conjunction with its associated processor to classify object types within images. The CLIP

model, trained on a diverse range of images, demonstrated superior capabilities in recognizing and categorizing various object types (Conde & Adobe, 2021).

The process involved loading and preprocessing the image, after which the model assigned probabilities to each image based on its content. The probabilities were determined using a list comprehension approach, where the model's predictions were obtained for each specified object category. OpenAI's CLIP model classifies images into categories by leveraging its ability to understand the relationship between text and images (Yoann Lemesle et al., 2022). CLIP is trained to associate images and their textual descriptions in a way that allows it to generalize across a wide range of tasks without task-specific training. Here's a simplified explanation of how the CLIP model works for image classification:

- **Pre-training:** CLIP is pre-trained on a large dataset containing images and their associated textual descriptions. The model learns to embed both images and text into a shared representation space.
- **Contrastive Learning:** The pre-training involves a contrastive learning objective. The model is trained to bring the representation of a given image closer to its textual description and push away the representations of other images and their descriptions. This process enables CLIP to understand the semantic relationships between images and text.
- **Generalization:** Once pre-trained, CLIP can generalize its understanding of images and text to perform various downstream tasks without task-specific training. This includes tasks like image classification, object detection, and more.
- **Zero-Shot Learning:** CLIP is capable of zero-shot learning, meaning it can make predictions on tasks it has never seen during training. This is because it can relate the new task to the general relationships it learned between images and text during pre-training.

In the process of image classification using CLIP, when an image is provided, the CLIP model generates a unique representation for that image within a shared space where textual descriptions also exist. This representation is then compared with the representations of provided textual categories called labels. To accomplish this classification task, a Python program utilized a set of labels (reference texts) representing distinct object categories for classification purposes. These labels were applied to classify all images within our dataset. Following the application of these labels, we systematically extracted the probabilities associated with object classification for each image.

- Human
- Animal
- Nature
- Vehicle or Machine
- Other (None of above, but other)

Each label is assigned a probability based on the similarity between the image representation and the textual representations (Radford et al., 2021). For instance, in scenarios where an image encompasses multiple objects, such as humans, animals, and elements of nature, CLIP classifies the image into one of these categories by identifying the dominant representation.

The following Figure 3 explains the probability distribution of classified object categories for each image of CSR files resulted during the object classification task.

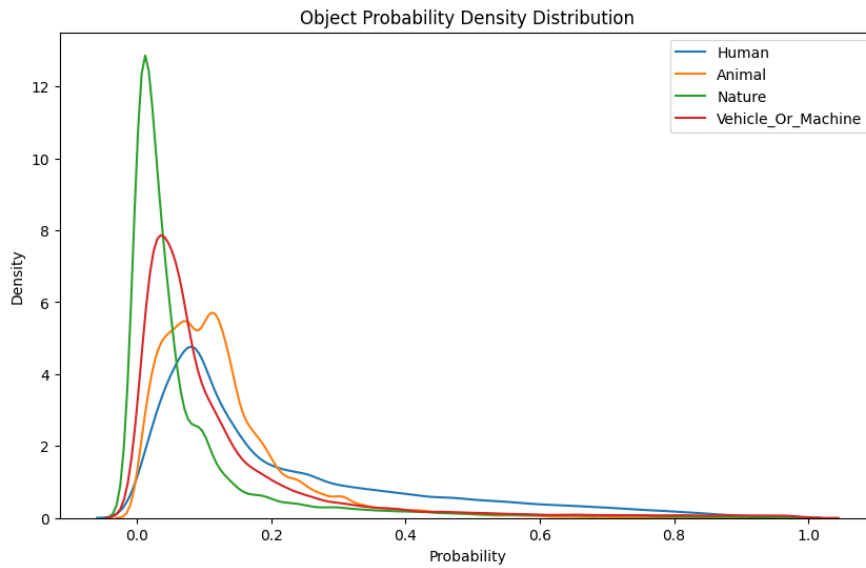


Figure 3: Density of classified object probabilities

The classified category is determined by selecting the one whose representation is most like the image representation. The associated probability for this predicted category is higher than those of other categories. In our analysis discussions, we use the term "dominant classified object" to denote the category with the highest probability and the probability value exceeding 0.5 (50%). Out of the 66,925 images we extracted from the filtered CSR files, we determined the dominant classified object for each image based on our defined criteria: probability > 0.5 . The Table 5 illustrates the count of each dominant classified object category derived from the image dataset.

Table 5: Classified Object Category

Dominant Classified Object Category	Count of Images	Percentage (%)
Other ('Other' or not included into either of dominant category)	53711	80.26
Human	7696	11.50
Vehicle_Or_Machine	2763	4.13
Nature	1903	2.84

Animal	852	1.27
--------	-----	------

Based on the above statistical analysis of classified object categories, a notable observation is the predominant presence of images categorized under 'Other'. The count attributed to 'Other' represents the dominant instances classified under this label (31418), and the images that do not fall within any specific dominant category (22293), which are characterized by classified object probability values less than 0.5 for each defined category in a particular image. Classified object category is selected by filtering the values that have a probability of greater than 0.5 for each category. We found that 22293 images have not recorded a probability less than 0.5 for all the categories. A manual review of 670 images in the random sample reveals a diverse range of content, including graphics, graphs, textual elements, and abstract artworks that do not align with the predefined labels of Human, Animal, Nature, or Vehicle or Machine. The substantial count within the 'Other' category suggests a significant portion of the dataset comprises images that defy easy classification within the specified object labels. While acknowledging this diversity, we are confident in obtaining a satisfactory count for Dominant Classified Object categories of Human, Animal, Nature, and Vehicle_Or_Machine, enabling us to proceed with our thesis analysis.

Distinguishing living objects through the OpenCV model and categorizing them as human or animal using OpenAI's CLIP model yield distinct outcomes. The OpenCV model produces a binary result, indicating the presence or absence of a living object in the image. In contrast, the CLIP model provides probabilities, offering a relevant categorization based on the overall visual content of the image. These two models operate independently, each leveraging its unique training approach and delivering different performance metrics.

Objects and colors in CSR images play a pivotal role in analysis as they serve as visual elements conveying significant messages. Objects, such as human figures, animals, or nature, can symbolize a company's commitment to social responsibility,

environmental sustainability, or specific causes. Colors, being powerful visual cues, contribute to the emotional tone of the image

By homing in on colour, object extraction and classification, we aim to delve deeper into the visual elements of CSR reports, enriching our understanding of the subtle yet influential aspects that contribute to the narrative presented in these visual representations.

3.4 Text Preprocessing

Text Cleaning for sentiment analysis and topic modelling involves several steps. First, the text data needs to be cleaned by removing irrelevant information such as special characters, numbers, and punctuation marks. Next, the text data is tokenized, meaning it is divided into individual words or tokens. After tokenization, the next step is to remove stop words, which are commonly used words that do not carry significant meaning in sentiment analysis such as "and", "the", "is", etc. Then, the text is subjected to case folding, which involves converting all characters to lowercase.

After case folding, the text data is ready for stemming or lemmatization. Stemming is a process that reduces words to their base or root form, while lemmatization aims to bring words to their dictionary or base form. Finally, after all the preprocessing steps are complete, the text data is ready to be fed into a sentiment analysis model for further classification and topic modelling. Natural Language Processing (NLP) with SpaCy is a popular open-source library in Python, which is known for its efficiency and ease of use in handling large volumes of text was used for text processing.

3.5 Sentiment Analysis

Sentiment analysis is a computational study that involves extracting and classifying hidden emotions, attitudes, opinions, moods, judgments, and feelings expressed by users in text data. It is a valuable technique for understanding public sentiment

towards various topics, products, or services. Sentiment analysis is a valuable technique for understanding public sentiment towards various topics, products, or services (Shah et al., 2018). The main purpose of sentiment analysis is to determine the polarity of text or documents, whether they are positive, negative, or neutral (Jamil, 2019). By monitoring and analysing content generated from users' activities on social media, sentiment analysis can reveal hidden patterns of public behaviour.

Image sentiment analysis plays a crucial role in understanding the overall sentiment and emotions conveyed through visual content in CSR reports. By analysing the sentiment surrounding the images, we can gain insights into how stakeholders perceive the company's social and environmental initiatives.

The sentiment analysis was performed for the extracted text around images and VADER sentiment tool was used to generate the sentiment score. The main reason to select this tool is because it is fast to implement and doesn't require any training data which make easy to use out-of-the-box. Also, it incorporates grammatical and syntactical rules to determining sentiment which helps to understand the context.

Based on the compound score generated by VADER sentiment tool the sentiment was classified as follows,

- Positive (compound ≥ 0.5)
- Neutral (compound > -0.5 and compound < 0.5)
- Negative (compound ≤ -0.5)

Figure 4 below represents the positive and negative sentiment distribution of the dataset by sector-wise. Density graphs shows the positive and negative sentiment distribution across the 3 sectors where Agri & Food sector represents high frequency of positive sentiment compared to other sectors.

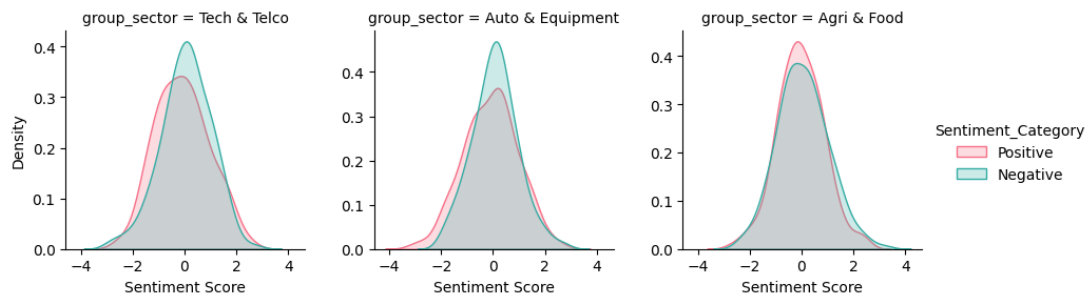


Figure 4: Density graphs of sentiment scores for positive and negative categories in sector groups

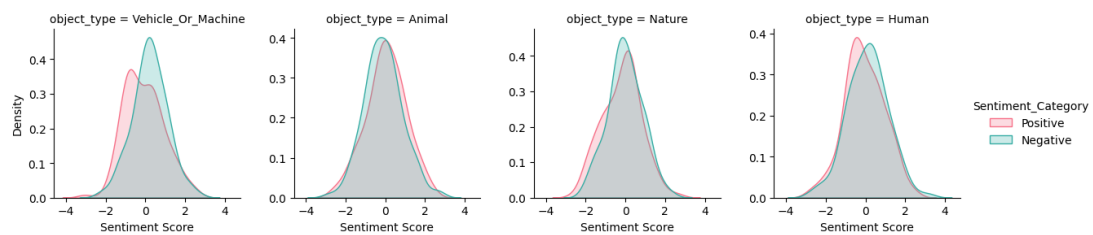


Figure 5: Density of sentiment scores for dominant objects

Figure 5 represents sentiment for each object type. From the four object types “Human” have the highest frequency of positive sentiment compared to others.

3.6 Topic Modelling

In the process of topic modelling, the text surrounding the images was considered. The implemented topic modelling function analyses the text in the 'Text_Before', 'Text_After', and 'Text_Inside_Image' columns of the data frame. It tallies the occurrences of words from each of the below-mentioned multi-dimensional corporate social responsibility words (CSRW) dictionary with four dimensions; Social and Community, Human Resources, Environment, and Human Rights developed by Nadra Pencle and Irina Malaescu (Pencle and Malaescu, 2016):

- CSRW - Employee: 319 words
- CSRW - Social and Community: 362 words
- CSRW - Human Rights: 297 words
- CSRW - Environment: 452 words

Topic modelling function implemented take two arguments text to analyse and a dictionaries object that contains multiple CSR-related dictionaries. The function begins by converting the text to lowercase to ensure consistent matching since dictionary words are also in lowercase. It then tokenizes the text into individual words by splitting the string on spaces. `topic_counts` is a dictionary comprehension that creates a dictionary with keys as the topics and values initialized to 0. This will store the count of occurrences for each CSR topic. The code iterates through each token in the tokenized text. For each token, it iterates through each topic and its associated list of words in the dictionaries. If the token is found in the list of words for a topic, the count for that topic in `topic_counts` is incremented by one. Then the topic counts are stored in separate field in dataset which is described in Table 3.

3.7 Data Indexing and Integration

After extracting colour, object, and text attributes, we proceeded to consolidate the final dataset by indexing and integrating image attributes with the GRI database. In this procedure, each image received a unique name through the combination of the CSR file name, page number, and image number. The Table 6 below succinctly summarizes these extracted and defined attributes, constituting a critical component of our research. It illustrates the essential data elements acquired through this amalgamation.

Table 6: Extracted and defined attributes for the analysis

Extraction	Attribute	Meaning	Data Type
Colour Extraction	Green Percentage	Percentage of Green Colour in the image	Numeric (%)

	Blue Percentage	Percentage of Blue Colour in the image	Numeric (%)
	Red Percentage	Percentage of Red Colour in the image	Numeric (%)
Object Extraction using OpenCV	Living Object	Whether Living Object (Human or Animal) is existed in the image	Boolean
Image Classification Using OpenAI CLIP model	Probability of Human Classified	These values are derived from the classification we did using OpenAI pre-trained model. These values indicate the probability that we can classify the image for each of these object types.	Numeric (0 to1)
	Probability of Animal Classified		Numeric (0 to1)
	Probability of Nature Classified		Numeric (0 to1)
	Probability of Vehicle or Machine Classified		Numeric (0 to1)
	Probability of Other Classified		Numeric (0 to1)
Sentiment	Sentiment Score	Sentiment Score of the text around and within the image that we extracted for different layers of the image	Numeric (-3 to 3)
Topic Modelling	CSRW - Employee	Topic count of the images extracted using CSRW (Corporate Social Responsibility Words) libraries for each topic	Numeric
	CSRW - Environment		Numeric
	CSRW - Human Rights		Numeric
	CSRW - Social and Community		Numeric
Grouping	Sector Group	Introduced Sector Groups: Agri & Food, Auto & Equipment, Tech & Telco	String

3.8 Exploratory Data Analysis (EDA)

The one of key objectives of this thesis is to conduct a comprehensive analysis of extracted attributes from both images and associated text. The aim is to discern relationships, contrasts, similarities, and trends, particularly concerning key company attributes; size, sector, and region. The analysis seeks to uncover hidden insights into the interplay between visual and textual elements in non-financial reports, shedding light on the potential influence of these attributes on stakeholder perceptions. As detailed in the 'Data Indexing and Integration' sub-sections of the methodology, numerous attributes were extracted in addition to the existing data fields of the original dataset. These attributes, derived from both images and text, play a pivotal role in shaping our analysis.

Before delving into the analysis, we formulated a strategic plan represented by Figure 6 in the form of a mind map. According to this mind map, we identified two primary elements: image and text. Images are broken down into two parts that are colours and objects. Text is broken down into two parts Topics and Sentiment.

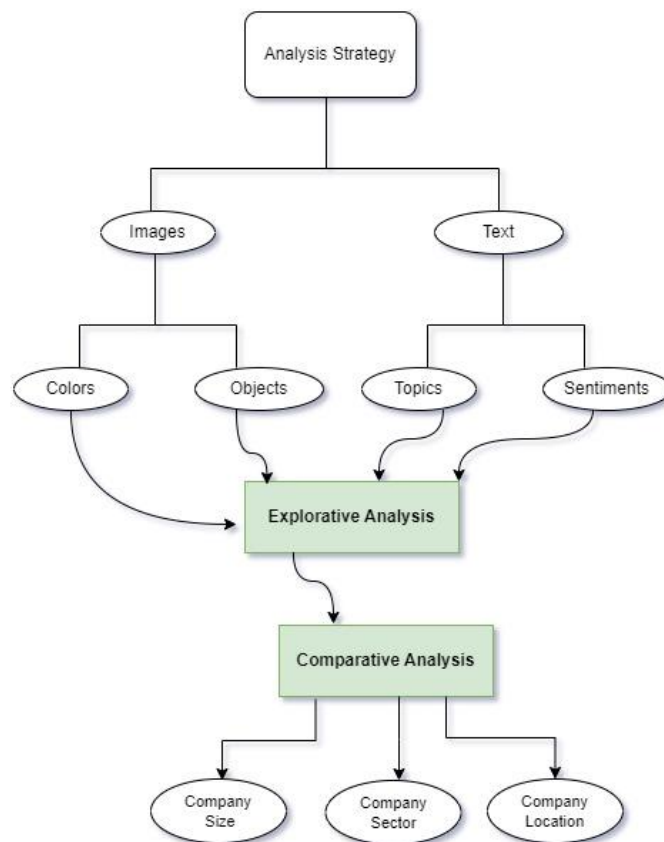


Figure 6: Analysis strategy of extracted data

1. Images – Extracted images from CSR reports
 - a. Colours – Green, Blue and Red colour percentages in each image
 - b. Objects – Living objects, and classified objects categories (probability values)
2. Texts – Extracted text associated with images
 - a. Topics: Analysing the subject matter or themes covered in the text.
 - b. Sentiments: Assessing the tone and content of the text to understand the sentiment it conveys.

The foundation of our final dataset is formed with extracted images, where each row corresponds to one image from a specific CSR report. Text attributes are associated with their respective images. The companies in our dataset are belongs to three size types: Large Enterprise (Large), Multinational Enterprises (MNE), and Small &

Medium Enterprises (SME). The dataset encompasses six regions ensuring the global representation of the CRS report sources. In our thesis work, we deliberately limited our focus to six sectors as explained in section 3.1. This intentional selection enhances the contrast in our analysis and augments, and the interpretability of our research outcomes. Additionally, we introduce group sectors as outlined in Table 7. Grouping sectors in the analysis was introduced for a more focused and insightful examination of CSR imagery. The sector groups, namely "Agri & Food," "Auto & Equipment," and "Tech and Telco," provide a structured approach to understanding how companies within specific industries present their CSR content. By grouping related sectors, we can identify common trends and unique characteristics within each group. This grouping allows for a more methodical exploration of sector-specific visual elements, such as the dominant objects and colors used in CSR images. It enables us to draw meaningful comparisons and contrasts between companies in similar industries, unveiling patterns that might be specific to certain sectors and enhancing the overall depth of our analysis.

Table 7: Sector Groups

Sector Group	Included Sectors
Agri & Food	Agriculture Food and Beverage Products
Auto & Equipment	Automotive Equipment
Tech and Telco	Technology Hardware Telecommunication

In the process of conveying our findings, we employed various visualization techniques to enhance clarity and comprehension. Utilizing Python's Matplotlib and Seaborn packages, we crafted an array of visual representations, including bar charts, scatter plots, density probability graphs, and line charts visualize the research findings. These visualizations serve as powerful tools to communicate complex insights in an accessible and visually appealing manner.

All the programs described in the above subsections in the methodology were executed on a personal computer with the specifications outlined in Table 8, showcasing the practical feasibility of our approach within a standard computing environment.

Table 8: Configuration of the personal computer program executed

Component	Configuration
Operating System	Windows 11 (64 bit)
Processor	13th Gen Intel(R) Core(TM) i5-1335U 1.30 GHz
RAM	16 GB
Hard Disc	SSD

As per our research objectives, we aimed at exploring relationships and trends between image and text attributes and comparison among company attributes such as size, sector, and Region. In the exploratory analysis, we scrutinized the interplay between image and text attributes to discern patterns between the visual components of images and their corresponding textual content. Our approach involved filtering the dataset based on the dominant classified object categories and primary topics. We delved into the analysis of colour distribution and sentiment across various dimensions. Concurrently, we seamlessly integrated comparative analysis with the exploratory study. This involved visualizing the outcomes while grouping companies based on sector, size, and region. The synergy between exploratory and comparative analyses proved instrumental in realizing the core objectives of our thesis.

3.9 Manual Sample Verification

Furthermore, to validate the performance of our machine learning models in living object identification and object classification, we conducted a manual verification process. This involved an in-depth comparison between the results generated by the models and those obtained through human visual assessment. Sample images of the

manual verification are listed in Appendix 1. We randomly selected 670 images, which constitutes approximately 1% of our extensive dataset for the manual verification. The manual verification process adds a layer of assurance to the reliability and accuracy of our systematic approach.

In the manual verification process, we assessed the performance of our system in two key aspects: the deviation in living object identification and the deviation in object classification. The OpenCV model returns whether there is a living object, while the CLIP model classifies images into object categories, based on provided text labels. There were instances where the OpenCV and CLIP models contradicted the results for human and animal objects. These results are dependent on the image quality (resolution) and other object types in the same image. Results of the manual sample verification are elaborated in the results section and, accuracy and performance matrices of each model are discussed.

4 Results

In pursuit of our objective to establish a systematic approach for extracting image attributes and associated text, we successfully developed a comprehensive framework explained in the methodology. This framework leverages machine learning models to extract various elements, including image objects, text sentiments, topics, and image classifications based on provided text labels. Our systematic approach was applied to a substantial dataset comprising around 66,925 images from 675 CSR reports. The key outcomes of this method, including the extraction of image attributes and text-related insights, represent significant achievements in the context of our thesis. Initially, the program's execution time for the entire dataset surpassed 7 days. However, through iterative optimization efforts, we successfully reduced the execution time to an impressive 27 hours. This achievement reflects our commitment to refining the code and enhancing the efficiency of our systematic approach.

Manual sample verification step of our systematic framework assures the reliability of our data mining methodology. Table 9 presents the summary of the manual verification results for the selected subset of 670 images.

Table 9: Summary of manual verification results

Verification Aspect	Percentage
Accuracy of living object identification by OpenCV model	81.19%
Accuracy of object classification (For all the objects) by CLIP model	76.42%
Accuracy of Human object classification	87.64%
Accuracy Human and Animal object Classification when it is failed to identify living objects (How CLIP model outperformed the OpenCV model)	96.87%

The results indicate an 81.19 accuracy in the identification of living objects and a 76.42% accuracy in the classification of objects compared to manual verification. As per our investigation of the sample verification, we found that CLIP model outperformed the OpenCV model in identifying human and animal objects. It is proven by the high accuracy of the CLIP model in correctly classifying human objects even when the living object identification process fails. The system demonstrated an impressive 96.87% accuracy in classifying human objects under these circumstances. This result highlights the robustness and effectiveness of the CLIP model in image classification, showcasing its ability to accurately identify human & animal related content in CSR images, even in cases where OpenCV might face challenges in detecting living object. The high percentage suggests that the CLIP model contributes significantly to the accuracy of human object classification, providing valuable insights for our CSR image analysis.

Furthermore, we conducted an in-depth examination of the outcomes, focusing on the maximum classified probability for each image in the presence or absence of a

Living Object. This analysis aimed to explore the relationship between the categorization of 'Human' and 'Animal' objects and the presence of Living Objects. The Table 10 presents the statistics for each classified object category with the recorded maximum value in each image, distinguishing between datasets where Living Objects exist and those where they do not, captured at the manual verification process. Remarkably, we observed an increase in the probability of 'Human' and 'Animal' objects in images containing Living Objects compared to images without living objects, highlighting the impact of the CLIP model on object classification based on the presence of living elements.

Table 10: Comparison of max probability derived from CLIP model for Objects categories with presence of Living objects

For images with Living Objects			For images without Living Objects		
Max Probability Object Category	Image Count	Percentage (%)	Max Probability Object Category	Image Count	Percentage (%)
Human	5994	47.08	Human	7163	13.22
Other	5024	39.46	Other	37727	69.62
Vehicle_Or_Machine	809	6.35	Vehicle_Or_Machine	3693	6.81
Nature	616	4.84	Nature	2994	5.52
Animal	289	2.27	Animal	2616	4.83
Total	12732	100	Total	54193	100

It's important to note that images in our dataset may contain multiple objects or no object at all. When considering the image object criteria, it's noteworthy that the CLIP model, used for classification, does not rely on predefined object categories or labels during its pre-training phase. Instead, CLIP learns to associate images with natural language descriptions, allowing for effective cross-modal retrieval. The model undergoes training on a diverse dataset encompassing a wide range of images

and their associated textual descriptions (Yang et al, 2023.) As a result, the CLIP model seeks to classify images by providing probability values that capture the overall essence of the images, rather than focusing on specific predefined object categories. This approach allows the model to generalize effectively and recognize diverse content, contributing to its versatility in image classification tasks.

Following the systematic execution of our methodology to capture image and text attributes, we translated our dataset into a series of visual representations aligning with our analysis strategy detailed in the preceding section. For reader convenience, we have organized only the key insights into three distinct subsections outlined below. Comparative analysis is seamlessly integrated into various visualization results obtained from exploratory analysis to enhance the effectiveness of communication. Some of dominant topic and dominant colour related findings are included into Appendix 2.

- Sector-based Findings
- Size-based Findings
- Region-based Findings

In our analysis, we have excluded the 'Other' object category from the dominant classified object visualizations due to its interpretational challenges when compared to the other object types. The dominant classified object is the object category where the probability derived from the object classification step is greater than 0.5. Similarly, the dominant colour is determined as the maximum colour percentage recorded for each image in the dataset.

4.1 Sector based findings

Figure 7 shows that in all sector groups Human is commonly used, when compared to other dominant object categories across sectors. We could observe nature percentage has recorded a significant growth in Agri & Food sector group compared to other sector groups, possibly to emphasize sustainability or origin of food

products. The dominance of Vehicle_Or_Machine objects in the Automotive & Equipment sector reflects the sector's focus on machinery and vehicles.

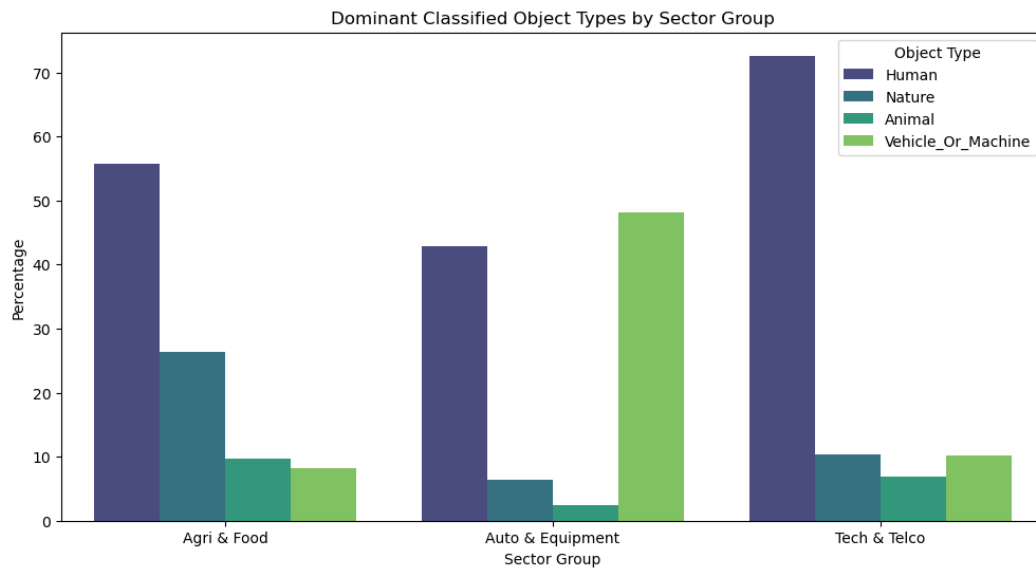


Figure 7: Dominant classified object type percentage in sector groups

Next, we conducted a time series analysis to assess the prevalence of various object types in CSR reports across different industries. Figure 8 illustrates the trends in image object categories within each sector over thirteen years, from 2006 to 2018. A notable dominance of 'Human' objects was observed in the 'Agri & Food' and 'Tech and Telco' sectors from 2010 to 2017. Conversely, from 2011 to 2014, the 'Auto & Equipment' sector experienced a marked decrease in 'Human' object depiction with a simultaneous rise in 'Vehicle_Or_Machine' imagery, indicating a shift in CSR focus toward vehicles and machinery. Additionally, the 'Agri & Food' sector consistently utilized 'Nature' objects more than the other sectors.

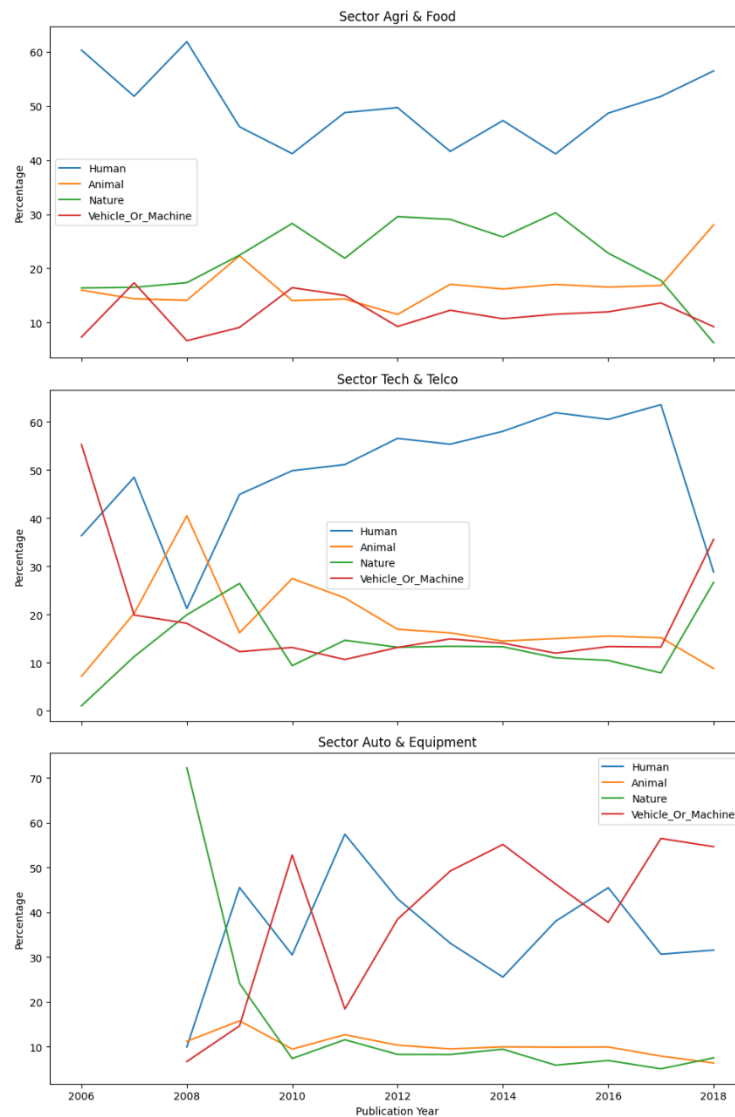


Figure 8: Time series analysis for dominant object percentage in sector groups

Looking at Figure 9 below, we can clearly see that 'Human' category is consistently the most represented in dominant CSR topics across all sector groups, highlighting the human-centric approach to CSR reporting. 'Nature' images are prominently featured in the Environment topic, indicating an emphasis on environmental aspects within CSR communications. This indicates the positive correlation between environmental topics and 'Nature' images over the other topics.

'Vehicle_Or_Machine' images are more frequently represented in the Auto & Equipment sector, which aligns with the industry's focus. The 'Animal' category is the least represented across all topics and sectors, suggesting it is not a focal point in CSR reporting. There is variability in how each sector represents different image object categories, which likely reflects the unique focus areas and values of each industry.

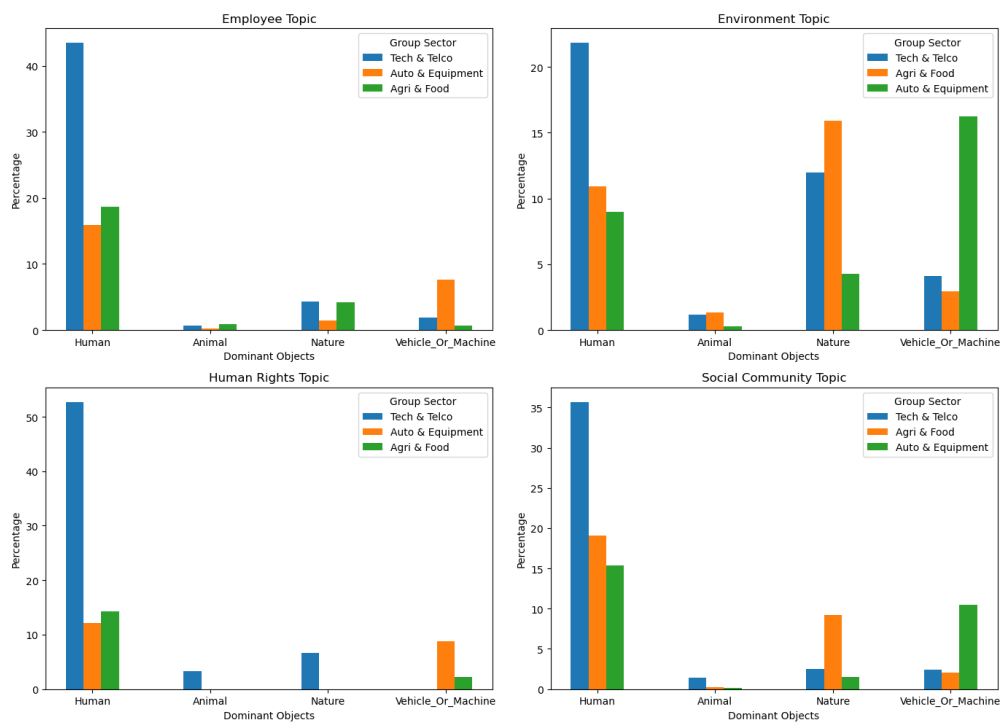


Figure 9: Dominant objects distribution among dominant topics for sector groups

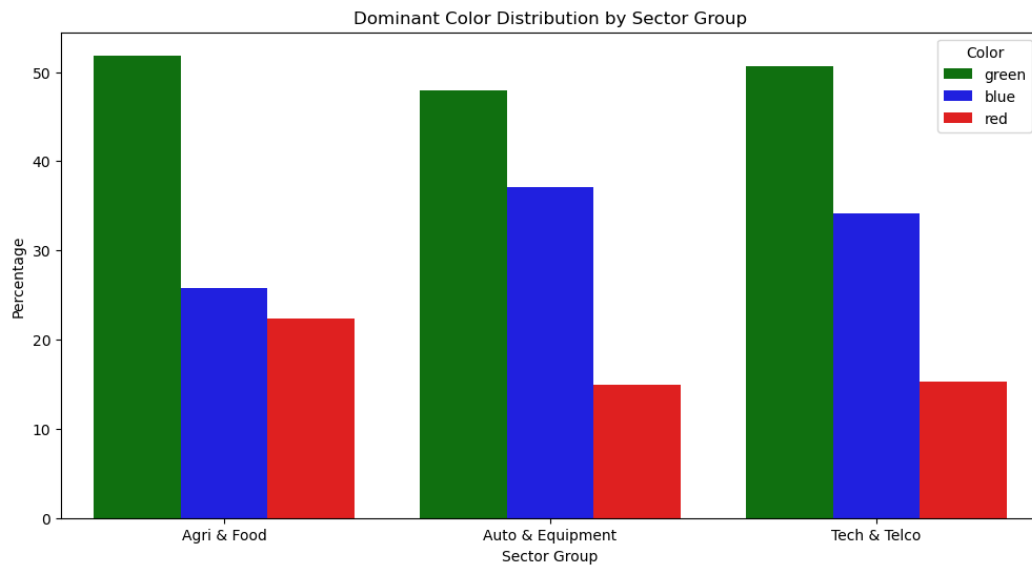


Figure 10: Dominant colour distribution in sector Groups

Figure 10 also shows that green is the most dominant colour across all sectors, which could indicate a general tendency to associate CSR with environmental themes or the colour's positive connotations. Blue is generally the second most common colour, potentially reflecting corporate and technological imagery. Red, while the least dominant overall, is relatively more present in the Automotive & Equipment and Tech & Telco sectors compared to Agriculture & Food. As per our analysis on region wise colour distribution, green and blue are dominant globally, while red is less prevalent.

Subsequently, we examined the behaviour of the sentiment scores in association with other attributes of the dataset to comprehend the impact of text sentiments on the disclosure of images.

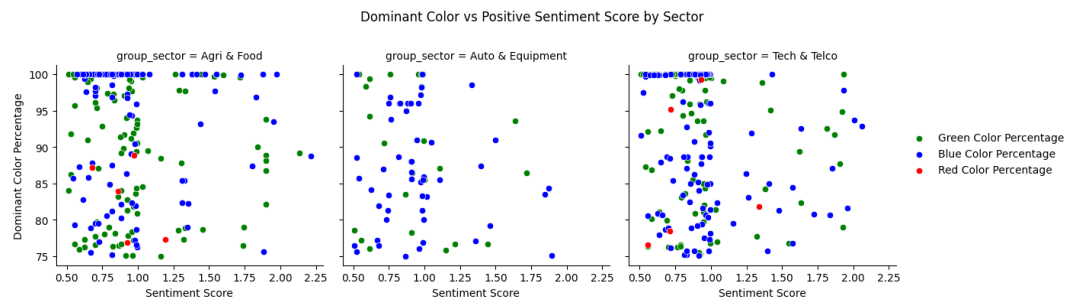


Figure 11: Correlation between sentiment and dominant colour in sector groups

Scatter plots composed in the Figure 11, represent the relationship between dominant colour percentages and positive sentiment scores across three different sector groups. The green colour percentage appears to be high in Agriculture & Food compared to other sectors. Blue is notable in the Automotive & Equipment and Tech & Telco sectors, likely reflecting technological and industrial themes. Red is consistently the least dominant colour across all sectors. The sentiment score does not show a consistent pattern in relation to colour percentages, indicating that the sentiment of the text associated with images may not be directly related to the dominant colour in the imagery.

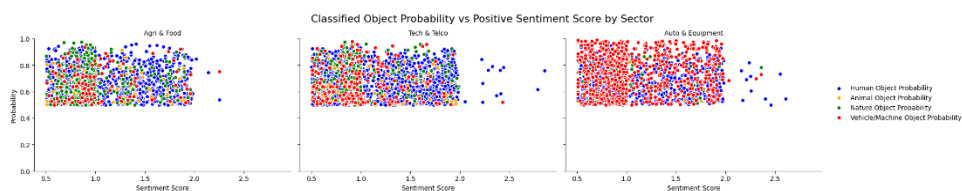


Figure 12: Correlation between classified object probability and positive sentiment in sector groups

The scatter plots included in the Figure 12 illustrate the relationship between the probability of classified objects and positive sentiment scores within sector groups. The positive sentiment score between 0.5 and 2.0 it clearly shows Vehicle/Machine probability is high in “Auto & Equipment” Sector group. Also, it shows that Human

object dominance in both “Agri & Food” & “Tech & Telco” Sector groups. Also, Nature probability is high in “Agri & Food” compared to other Sector groups.

The scatter plots in the Figure 13 compare the probability of various classified objects against negative sentiment scores within the sector groups. Compared to the behaviour of positive sentiment across the sectors there is no significant pattern observed in negative sentiment in relationship with classified object probability.

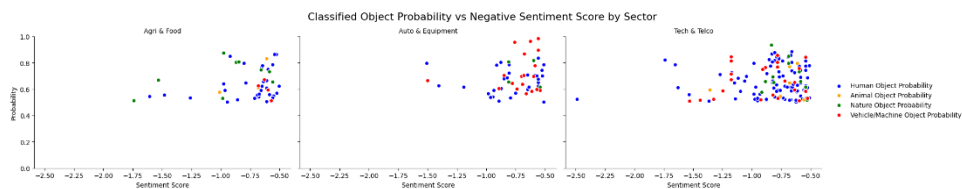


Figure 13: Correlation of classified object probability and negative sentiment in sector groups

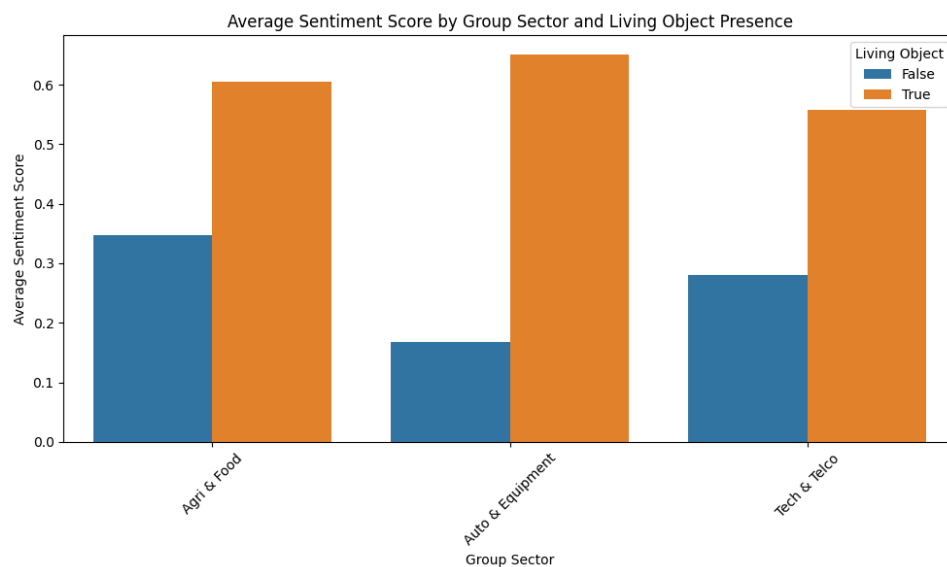


Figure 14: Average sentiment score distribution with presence of living object in sector groups

The average sentiment scores for each sector group are visualized, considering images with both living objects present and absent. The findings in Figure 14 suggest that the presence of living objects in images could be correlated with a higher average sentiment score in all three sectors, indicating a more positive response when living objects are featured. In the Agri & Food sector appears to have a higher average sentiment score when living objects are absent from images and the second-best average sentiment score with comparison to other sectors when living object is present.

4.2 Company Size based findings

With the interest of exploratory analysis, our initial step involved visualizing the distribution of dominant object percentages based on company size.

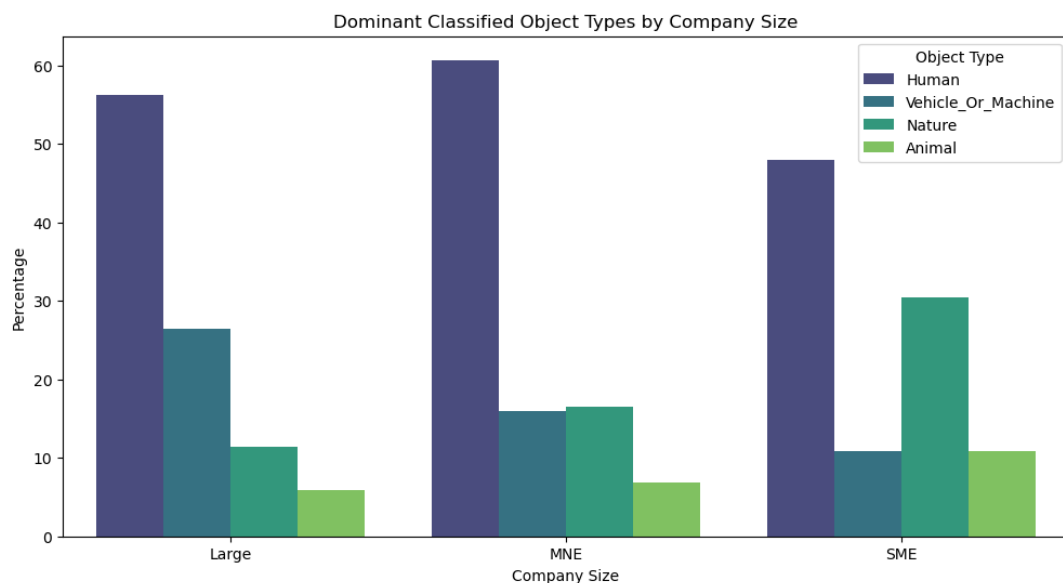


Figure 15: Dominant classified object type percentage in company size

In the Figure 15 it shows that across all company sizes, animal objects are the least represented, indicating that animal imagery is not a primary focus in CSR reporting regardless of the company size. Human is the more commonly represented in CSR reporting, similar to sector-wise findings. Nature object representation is

considerably high in SME compared to LARGE and MNE and Vehicle_Or_Machine object representation is the highest in LARGE companies compared to other company sizes.

As indicated by Figure 16, multinational enterprises (MNEs) consistently exhibit the highest utilization of human objects across various topics used in the text around each image, except for one instance where large companies surpass them in the social and community topic. MNEs predominantly feature 'Nature' objects in all topics, apart from the Human Rights topic. The utilization of 'Vehicle_Or_Machine' objects is predominantly dominated by large companies across all topics, excluding the Human Rights topic.

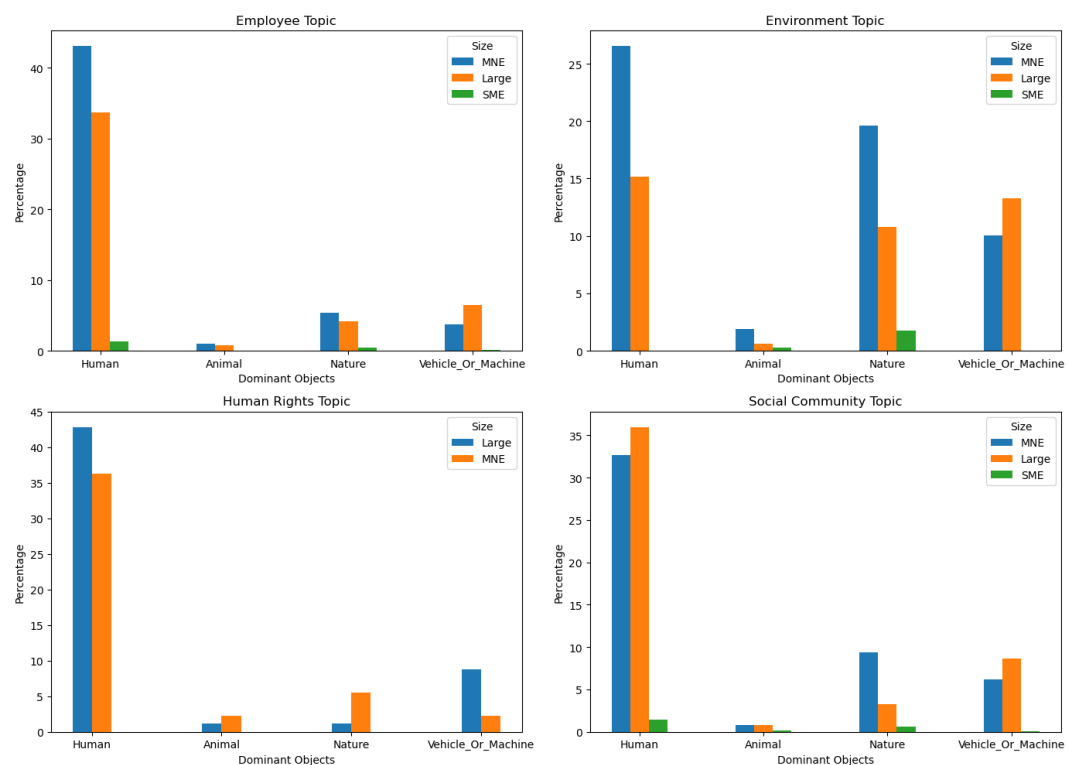


Figure 16: Dominant objects and topic distribution in company size

4.3 Company Region based findings

As illustrated in Figure 17, the human objects are consistently among the most represented across the regions. Vehicle_Or_Machine objects are more variably

represented, likely reflecting regional industrial differences. Animal objects are consistently the least represented across all regions, indicating that they are less frequently used or are not the focus of CSR imagery in these regions. Nature Objects are widely used in Africa and Oceania regions compared to other regions.

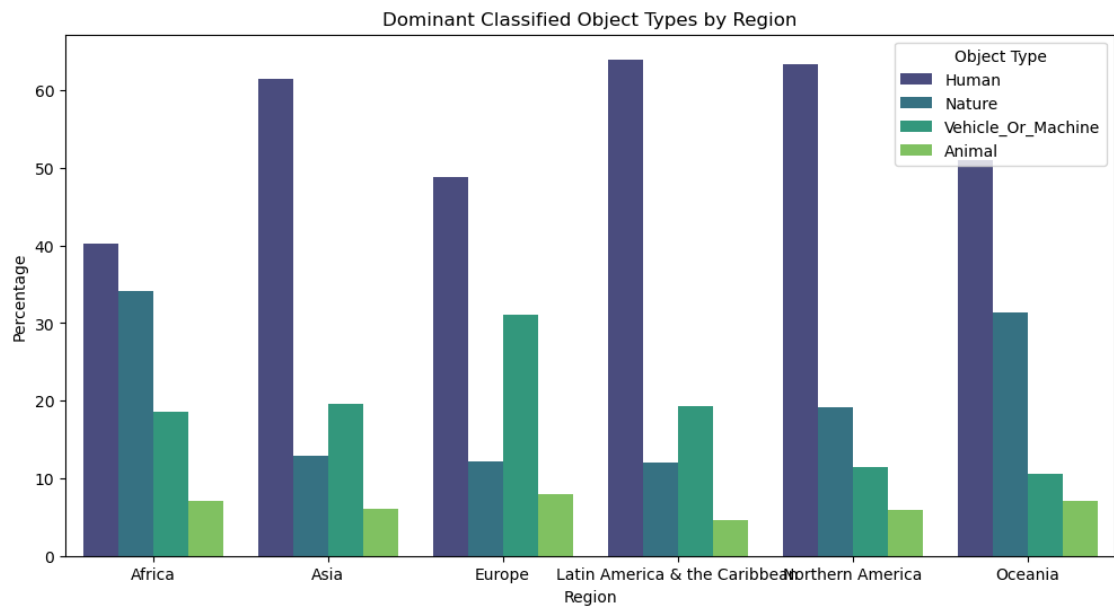
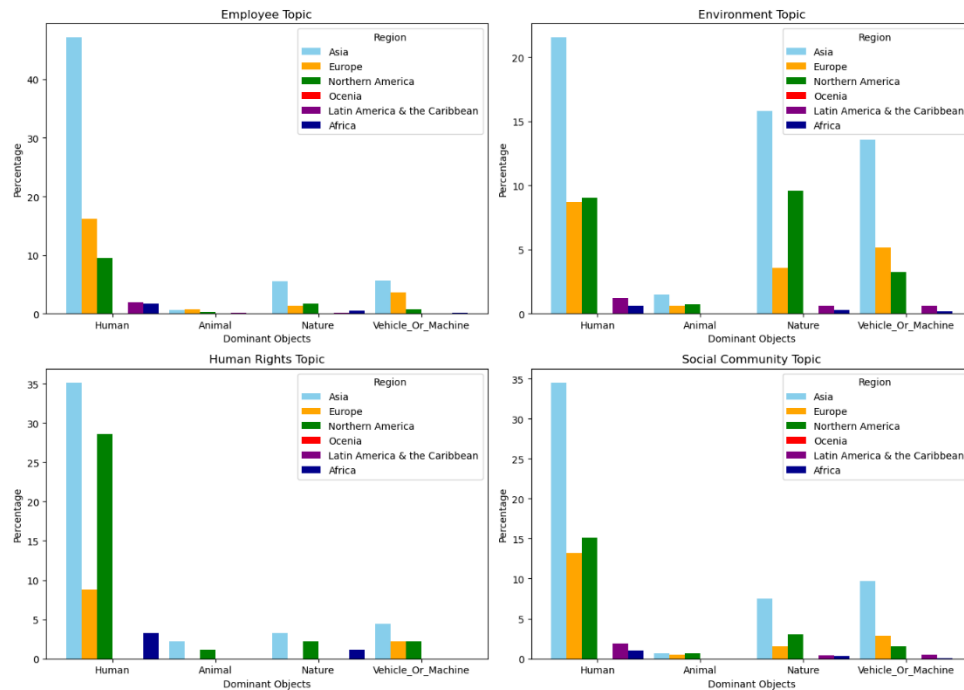


Figure 17: Dominant classified object type percentage in regions

In the Figure 18, across all topics and regions, the Human category is consistently the most associated with the given CSR topics, which is logical considering the social nature of these topics. The Nature category is predominantly associated with the Environment topic, particularly in Asia and Africa. The Vehicle_Or_Machine category, while generally lower, shows a unique spike in the Environment topic for Asia, suggesting a regional emphasis on mechanization or industry in environmental discussions. Animal-related content is the least associated with CSR topics across all regions.



Figure

18: Dominant objects and topic distribution in region

Then we wanted to analyse the relationship between average Sentiment Score and Living Object and compare the regional results. There is a clear trend observed in figure 19, when there is Living object present in the image, the average sentiment score is high in all the regions. This confirms most organization uses human and animal objects (living objects) in the images used in CSR reports to convey positive message.

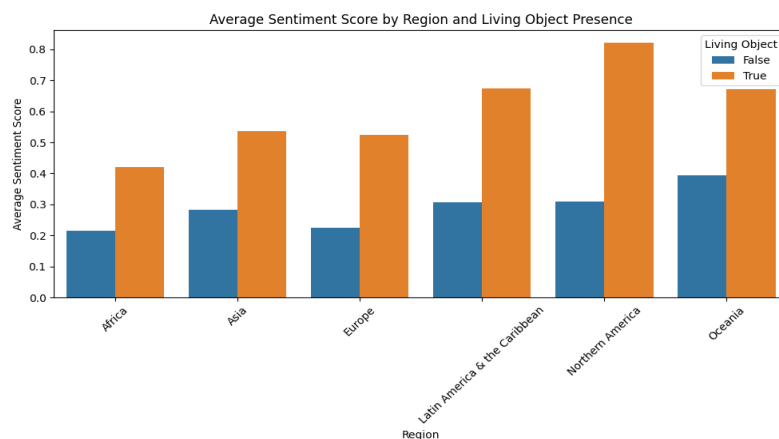


Figure 19: Distribution of average sentiment score with presence of living objects in regions

Top keywords categorized into positive, negative, and neutral sentiments are listed in Table 2 at appendix 2.

5 Discussion

Our results reveal a distinct interplay between image and text attributes in CSR communications. While certain trends are consistent globally, regional, and sector-specific variations highlight the strategic use of visual elements to align with diverse communication goals. These findings contribute valuable insights into the dynamic relationship between image attributes and textual content in CSR reports across diverse organizational contexts. We would like to highlight key observations derived from our thesis work, aligning with the three objectives of our research.

The methodology of our research aimed to develop and implement a systematic approach for extracting image attributes and associated text, utilizing advanced data mining techniques. The systematic framework we developed consist with comprehensive image and text extraction process using OCR, object and color extraction, object classification, sentiment analysis, topic modelling and exploratory data analysis steps demonstrated in Figure 1. By employing this approach, we successfully captured a rich dataset, allowing for a comprehensive analysis of the visual and textual discourse in CSR reports. The systematic extraction process laid the foundation for generating valuable insights, enhancing the efficiency and reliability of our corporate sustainability analysis. Kexin has successfully classified satellite land cover images using CLIP model which led to a 2.5% improvement in zero-shot classification accuracy on the AID dataset (Meng, 2023). Similar research conducted by Karanam aimed to establish an image processing framework capable of rapidly and accurately classifying bone fractures using data from X-ray and Computer Tomography (CT) scans (Karanam et al., 2022).

In addressing the objective of investigating the utilization of image attributes in CSR reports, we delved into the relationship between these attributes and key company characteristics, including size, location, and sector. Our findings revealed intriguing

patterns, indicating that certain image attributes are strategically employed in alignment with the unique communication goals and values of companies. The exploratory analysis provides a deep understanding of how visual elements contribute to the overall CSR narrative, offering implications for corporate branding and stakeholder engagement. Our exploration of sentiments and topics within CSR reports uncovered the complex connections with the attributes of featured images. By scrutinizing the sentiment and thematic content, we gained insights into the emotive and informational dimensions of CSR communications. The analysis exposed trends in the usage of image attributes across different topics, shedding light on the strategic deployment of visual elements to reinforce specific CSR themes. This thorough understanding contributes to a deeper comprehension of the communicative strategies employed by organizations to convey their commitment to sustainability.

5.1 Limitations of the Study

Our study has uncovered important insights, yet we must recognize certain limitations that could affect the extent to which our conclusions can be applied. The dataset's concentration on sectors such as agriculture, food, automotive, and technology may not translate seamlessly to different industries. Additionally, the use of the CLIP model for categorizing images leans on the model's existing knowledge, which might not capture unique industry-specific subtleties. The timespan of the dataset, which covers 2006 to 2018, might not reflect the most current trends in CSR communication.

In extracting text adjacent to images, we selected ten lines preceding and following each image; however, we could not verify the direct relevance of this text to the image. This indicates a potential gap in our research approach, as the text analysed might not always correspond to the image in question.

A significant hurdle we faced was identifying and omitting minor images within the CSR reports that, despite their small size, could carry little pertinent information for

the text they accompany. Our method does not have the capability to discern and eliminate such images effectively. Despite setting a filter to exclude pages with a high image count (over ten), we cannot guarantee that all non-essential images were excluded. Moreover, certain image formats were not compatible with our system, leading to their exclusion from the analysis and the possible loss of useful information. Our framework's design is tailored for English-language CSR reports, which narrows the study's applicability to documents only in English.

Finally, the framework categorizes images into four distinct groups: Human, Nature, Animal, and Vehicle or Machine, disregarding other object types that might hold relevance in CSR discussions. This limitation could mean that significant objects are not recognized or are misclassified, which is a notable constraint of our study.

5.2 Recommendations for Future Research

Future research should focus on refining methodologies that accurately determine the contextual relationship between text and its accompanying imagery. This may include the application of advanced natural language processing (NLP) tools or the development of innovative machine learning models adept at identifying semantic connections between text and images with enhanced accuracy (Khurana et al., 2022). One such approach could be the adoption of the visual-textual sentiment classification model known as Multi-Model Fusion (MMF), devised by K. Salman Al-Tameemi. This model employs a mixed fusion framework, designed to capture vital information and the inherent relationship between visual and textual elements. It consists of three separate deep neural networks, each specifically engineered to extract the most emotionally pertinent features from both image and text data, facilitating the collection of more distinctive features for accurate sentiment classification (K. Salman Al-Tameemi et al., 2023). Future studies should also prioritize the improvement of algorithms that assess the significance of images within CSR narratives, especially smaller ones. The goal would be to differentiate images that are decorative from those that significantly contribute to CSR communication narratives.

Further efforts should explore advanced techniques to facilitate the processing and interpretation of a wider array of image formats, which are not presently accommodated by the existing framework. This would entail the development of more sophisticated image processing tools that can adapt to diverse image types and avoid miscategorising them as corrupted. To widen the scope of the study, future research should look to adapt the framework for the analysis of CSR reports across a variety of languages. An et al developed Chinese text to be trained with CLIP model for large dataset successfully (An et al., 2022). This would involve integrating technologies adept at handling multilingual text, thereby making the framework more comprehensive and applicable to international CSR analyses.

To deepen the analysis of visual content in CSR reports, it is essential for forthcoming research to extend the classification taxonomy within the framework. By incorporating a wider array of object types, the framework would be better equipped to recognize and categorize a more diverse set of objects present in imagery, enriching the insights drawn from visual data in CSR contexts.

6 Conclusion

In conclusion, we successfully achieved our first research objective by developing a systematic framework using data mining techniques to extract, classify and analyse image attributes and associated text. This framework depicted by Figure 1 , despite the diverse nature of CSR report images, demonstrated high accuracy of 81.19% in identifying living objects with OpenCV and 76.42% in overall object classification using the OpenAI - CLIP model. The CLIP model surpassed the OpenCV model in human and animal object identification, showcasing remarkable accuracy of 96.87% in classifying human objects even when the living object identification process encountered challenges. Those looking to advance our research methodology can explore further by addressing its limitations in future endeavours.

In fulfilling our second objective, we analysed image attributes in CSR reports, revealing trends across sectors, company sizes, and regions. Human-centric images

prevailed in CSR communication. Agri & Food emphasized sustainability with a surge in nature images. Automotive & Equipment sector featured Vehicle_Or_Machine objects, aligning with industry focus. Company size analysis showed consistent human-centric representation but varied Vehicle_Or_Machine objects. Regionally, 'Human' objects were consistent, while 'Vehicle_Or_Machine' varied, reflecting industrial differences. Nature objects were prominent in Africa and Oceania. The animal is the least used object type in every context.

Analysing relationship of text attributes (sentiments and topics) with image attributes within CSR reports contributes to our third objective of the thesis. Human is the mostly used object in all four topics; Employee, Environment, Human Right, Social Community. This highlighted a human-centric approach to CSR reporting. Environmental topics prominently featured 'Nature' images, and 'Vehicle_Or_Machine' images were more prevalent in sectors aligned with their focus.

In summary, our findings provide a thorough understanding of the dynamic interplay between image attributes and textual content in CSR reports, shedding light on sector-specific, size-related, and region-specific variations. This research contributes valuable insights to the field and paves the way for more comprehensive analyses in the intersection of visual and textual elements in corporate sustainability communications.

References

- Akinbade, Daniel, et al. "An Adaptive Thresholding Algorithm-Based Optical Character Recognition System for Information Extraction in Complex Images." *Journal of Computer Science*, vol. 16, no. 6, 1 June 2020, pp. 784–801, <https://doi.org/10.3844/jcssp.2020.784.801>. Accessed 7 Oct. 2023.
- An, Y., Pan, J., Jung Fu Lin, Men, R., Zhang, Y., Zhou, J., & Zhou, C. (2022). Chinese CLIP: Contrastive Vision-Language Pretraining in Chinese. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2211.01335>
- Ashwini, Balaji, V., Srinivasan, S., & Kavya Monisha. (2020). Real Time Facial Recognition System. Springer eBooks, 1721–1726. https://doi.org/10.1007/978-3-030-41862-5_175
- Anantharaman, D., Huang, D., & Zhao, K. (2020). Is a Picture Worth a Thousand Words? Image Usage in CSR Reports. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3722228>
- Breitbarth, T., Harris, P., & Insch, A. (2010). Pictures at an exhibition revisited: reflections on a typology of images used in the construction of corporate social responsibility and sustainability in non-financial corporate reporting. *Journal of Public Affairs*, 10(4), 238-257.
- Buallay, Amina. "Sustainability Reporting in Food Industry: An Innovative Tool for Enhancing Financial Performance." *British Food Journal*, vol. ahead-of-print, no. ahead-of-print, 29 Oct. 2021, <https://doi.org/10.1108/bfj-01-2021-0053>.
- Cleary Gottlieb Steen & Hamilton LLP. (2023). The EU Corporate Sustainability Reporting Directive. Retrieved from <https://www.clearygottlieb.com>
- Cooley LLP. (2023). EU Changes CSRD Thresholds to Reduce Sustainability Reporting Burden. Retrieved from <https://www.cooley.com>

- Cortés-Sánchez, J.D., 2022. What is the mission of innovation?—Lexical structure, sentiment analysis, and cosine similarity of mission statements of research-knowledge intensive institutions
- Conde, M., & Adobe, K. (2021). CLIP-Art: Contrastive Pre-training for Fine-Grained Art Classification. https://openaccess.thecvf.com/content/CVPR2021W/CVFAD/papers/Conde_CLIP-Art_Contrastive_Pre-Training_for_Fine-Grained_Art_Classification_CVPRW_2021_paper.pdf
- Duraivel, S. & Lavanya, R., 2022. VOD and OTT Content as a Therapeutic Measure to Improve Self-esteem: A Retrospective Analysis of user Generated Opinion using NLP Techniques.
- European Commission. (2023). Corporate sustainability reporting. Retrieved from <https://finance.ec.europa.eu>
- GRI. (2022). Mission & history. [Www.globalreporting.org](http://www.globalreporting.org). <https://www.globalreporting.org/about-gri/mission-history/>
- Harvard Law School Forum on Corporate Governance. (2023). EU Finalizes ESG Reporting Rules with International Impacts. Retrieved from <https://corpgov.law.harvard.edu>
- Hasan, R. T., & Sallow, A. B. (2021). Face Detection and Recognition Using OpenCV. *Journal of Soft Computing and Data Mining*, 2(2), 86–97. <https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/8791>
- Hellmueller, L., & Zhang, X. (2019). Shifting toward a humanized perspective? Visual framing analysis of the coverage of refugees on CNN and Spiegel Online before and after the iconic photo publication of Alan Kurdi. *Visual Communication*, 1470357219832790.

- IBM. (2023). What is the Corporate Sustainability Reporting Directive (CSRD)? Retrieved from <https://www.ibm.com>
- Invernizzi, A C., Bellucci, M., Acuti, D., & Manetti, G. (2021, December 23). Form and substance: Visual content in CSR reports and investors' perceptions. <https://scite.ai/reports/10.1002/mar.21635>
- Jamil, N S. (2019, December 15). Social Tension and Crime Related Events Detection Method on Twitter. <https://scite.ai/reports/10.30534/ijatcse/2019/23862019>
- Jackson, Gregory, et al. "Mandatory Non-Financial Disclosure and Its Influence on CSR: An International Comparison." *Journal of Business Ethics*, 12 June 2019, <https://doi.org/10.1007/s10551-019-04200-0>.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Lindgren, C., Huq, A. M., & Carling, K. (2021). Who are the intended users of CSR reports? Insights from a data-driven approach. *Sustainability*, 13(3), 1070.
- Luo, X., Zhang, Y. & Zhou, J., 2023. A subject detection strategy combined with image categorization to classify objects in CSR reports. *Journal of CSR Research*, 15(2), pp.123-145.
- Meng, K. (2023). APPLYING CLIP FOR LAND COVER CLASSIFICATION USING AERIAL AND SATELLITE IMAGERY. Purdue University Libraries and School of Information Studies. <https://doi.org/10.25394/PGS.24717393.v1>

- Miethlich, B. (2019). Disability as a blind spot in sustainability frameworks and standards. *International Journal of Recent Technology and Engineering*, 8(4).
- Moses, L., 2021. Meaningful messaging: Sentiment in elite social media communication with the public on the COVID-19 pandemic.
- Mučko, P., 2021. Sentiment analysis of CSR disclosures in annual reports of EU companies
- Muttineni, S. & Deng, J., 2021. Estimating Sentiment in Social Media - a Case Study of the Migrant Caravans News on Twitter.
- Rebecca Stone. (2018, September 26). Image Segmentation Using Color Spaces in OpenCV + Python. *Realpython.com*; Real Python. <https://realpython.com/python-opencv-color-spaces/>
- Radford, A., Kim, J., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Aspell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. <https://proceedings.mlr.press/v139/radford21a/radford21a.pdf>
- Sung, K., Tao, C. & Slevitch, L., 2022. Do strategy and content matter? Restaurant firms' corporate social responsibility communication on Twitter: A social network theory perspective.
- Serafini, F., & Reid, S. F. (2023). Multimodal content analysis: expanding analytical approaches to content analysis. *Visual Communication*, 22(4), 623-649.
- Shah, D., Isah, H., & Zulkernine, F. (2018, December 1). Predicting the Effects of News Sentiments on the Stock Market. <https://scite.ai/reports/10.1109/bigdata.2018.8621884>

- Wang, X., Zhang, L., Wang, N., & Zou, B. (2021, January 1). Joint Polarimetric-Adjacent Features Based on LCSR for PolSAR Image Classification. <https://scite.ai/reports/10.1109/jstars.2021.3087164>
- Wickert, C. (2021). Corporate Social Responsibility Research in the Journal of Management Studies : A Shift from a Business-Centric to a Society-Centric Focus. *Journal of Management Studies*, 58(8). <https://doi.org/10.1111/joms.12775>
- Yang, Yue, et al. "Language in a Bottle: Language Model Guided Concept Bottlenecks for Interpretable Image Classification." *Openaccess.thecvf.com*, 2023, openaccess.thecvf.com/content/CVPR2023/html/Yang_Language_in_a_Bottle_Language_Model_Guided_Concept_Bottlenecks_for_CVPR_2023_paper.html. Accessed 27 Dec. 2023.
- Yoann Lemesle, Masataka Sawayama, Valle-Perez, G., Adolphe, M., H el ene Sauz eon, & Pierre-Yves Oudeyer. (2022). Language-biased image classification: evaluation based on semantic representations. ArXiv (Cornell University). <https://doi.org/10.48550/arxiv.2201.11014>

Appendix 1 – Results of Manual Verification on Object Identification and Object Classification

CSR Report Name - Qisda_2010

Image Name - page_50_image_2.png



Human image upside down, living object Identification failed, but classification success.

CSR Report - PiaggioGroup_2012

Image Name - page_36_image_1.png



Bike Design like a human face, Human object detected, but correctly classified as a Vehicle/Machine (0.9439)

CSR Report Name -TelenorHungary_2015

Image Name -page_25_image_1.png



Human Painting, Living Object Identification and Classification failed.

CSR Report Name - Thaibev_2016

Image Name - page_98_image_3.png



Image is not focused and low resolution, living object Identification failed, but classification success.

CSR Report -ToyotaGoseiCompanyLimited_2013 -

Image Name - page_25_image_1.png



Image is not focused and low resolution, living object Identification failed, and dominant object classification is also not given as human, but maximum probability is given as a Human (0.4811)

Appendix 2 – Dominant Topic and Dominant Colour based findings.

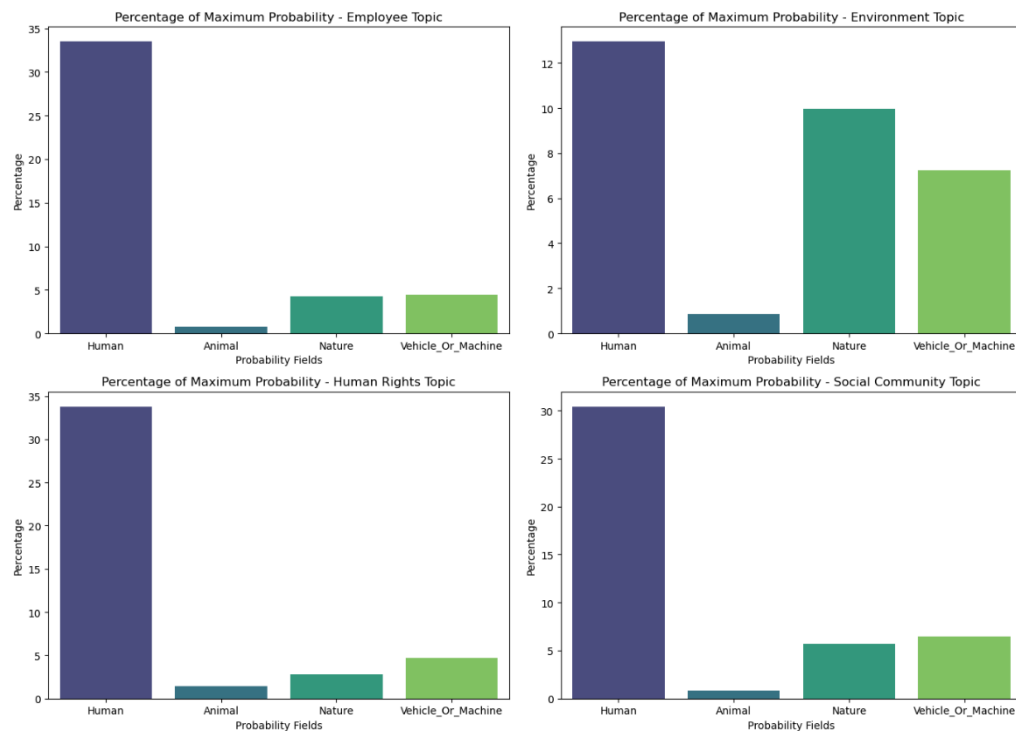


Figure 20: Dominant classified object percentages among dominant topic categories

When we investigate the correlation between dominant classified object and dominant topic categories, shows in the Figure 20, the 'Human' category is consistently the most represented in topics directly related to human-centric issues (Employee, Human Rights, and Social Community), indicating a strong visual emphasis on people in these areas of CSR communication. The 'Nature' category is prominently represented in the Environment Topic, which suggests that natural imagery is a key visual tool for communicating environmental concerns. The 'Animal' and 'Vehicle_Or_Machine' categories are less represented across all topics, implying that these elements are not the primary focus in the visual strategies for the CSR topics considered.

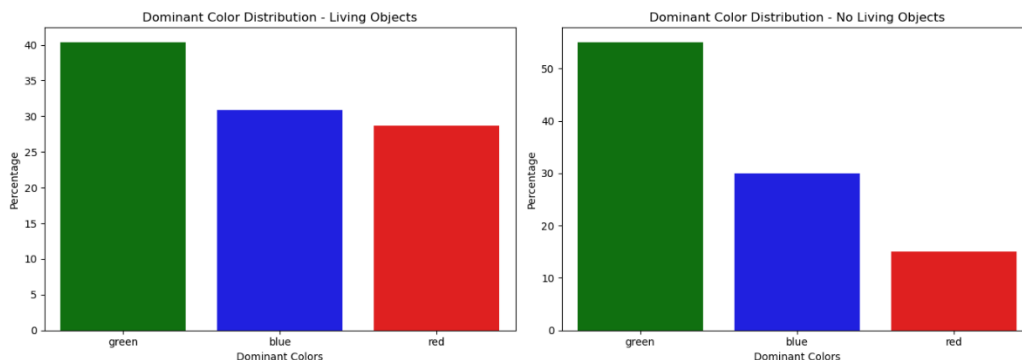


Figure 21: Dominant colour percentage with presence of living objects

We aimed to observe the contrast in dominant colour distribution between datasets where living objects exist and where they do not. As depicted in Figure 21, green emerges as the most dominant colour in both categories, indicating that environments, regardless of the presence of living objects, commonly exhibit green hues. Conversely, red consistently appears as the least dominant colour, possibly due to its less frequent natural occurrence in the depicted environments.

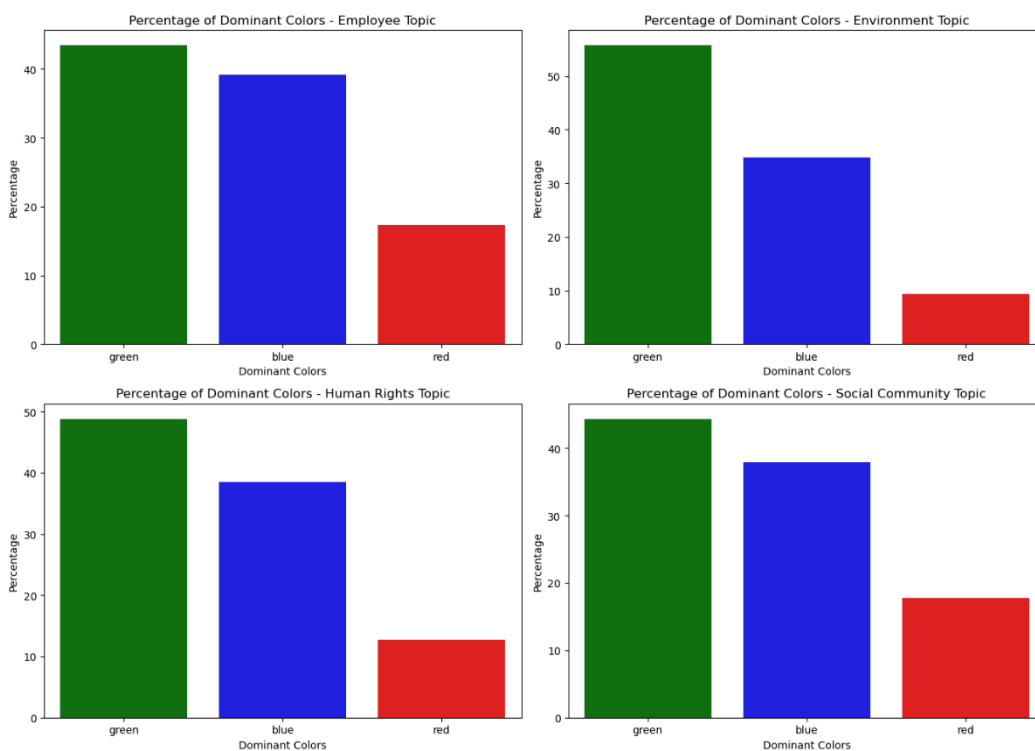


Figure 22: Dominant colour distribution in dominant topics

There is clear that trend observed in Figure 22, which indicates green and blue colour is widely used when discussing all 4 topics in CSR reports. In all four dominant CSR topics related images, green is most dominant colour, suggesting that it is a common colour used in CSR communications, possibly due to its positive connotations with growth, nature, and sustainability. Blue is also prominent, particularly in the Employee and Human Rights topics, which may be due to its association with professionalism and trustworthiness. Red is consistently the least dominant colour across all topics, which might be due to its strong association with urgency or danger, emotions that are typically not the focus of CSR communications.

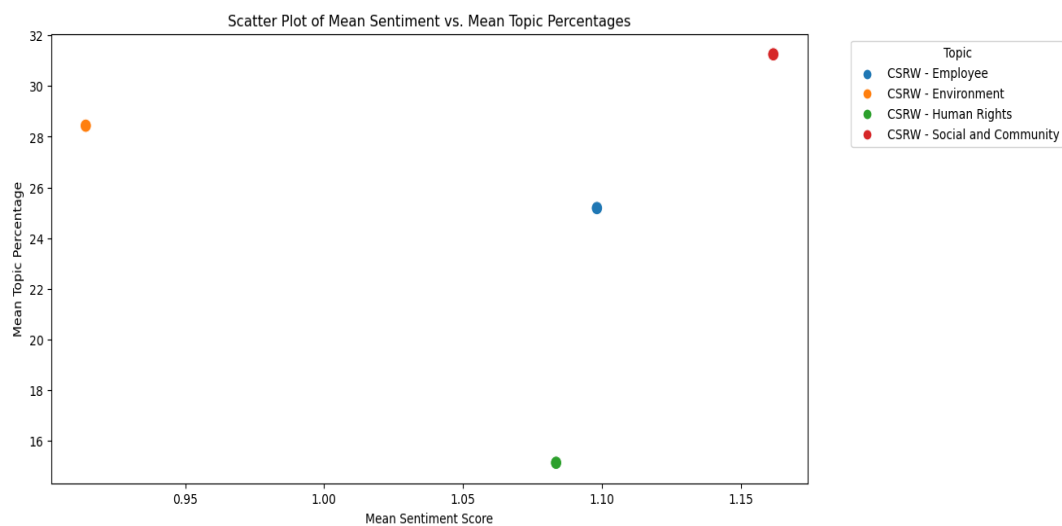


Figure 23: Correlation between Dominant Topic percentage and mean sentiment

It is intriguing to explore the correlation between dominant topic percentage and the mean sentiment for each topic in the dataset. This analysis provides insights into the prevalence of each topic across all images and the corresponding sentiment associated with each topic. The figure 23 illustrates a comparison between the average sentiment score and the mean percentage of four topics: Employee, Environment, Human Rights, and Social and Community. The data indicates that discussions pertaining to the Employee topic often carry the highest mean sentiment score, suggesting a generally positive tone in communications related to employees.

Conversely, the Environment topic tends to have a lower mean sentiment score on average, which could be attributed to discussions around potentially negative issues like pollution and environmental hazards.

Appendix 3 – Thesis work plan and individual contribution.

#	Task	Week Number (Week 1 to 20)	Contribution (%)	
			Julian	Nilupa
1	Background Discussion	1,2	50	50
2	Technical Feasibility Study	2,3,4	50	50
3	Project Planning	5	50	50
4	Dataset Examination	5,6	40	60
5	Dataset Criteria Selection (Scoping)	6	60	40
6	Image and text extraction	4,5,6	100	0
7	Framework Finalization (Methodology)	6,7	50	50
8	Living Object Extraction	7,8	100	0
9	Framework optimization	8,9,10	0	100
10	Generating Sentiments	8,9	0	100
11	Image Classification	8,9,10,11	100	0
12	Topic Modeling	11,12	0	100
13	Exploratory Data Analysis	13,14,15,16,17,18	50	50
14	Manual Verification	17,18	50	50
15	Report Writing	14,15,16,17,18,19,20	50	50