Research Article

Text mining on sustainability reports of top 40 airlines and bibliometric analysis of airline's sustainability

İbrahim ŞAPALOĞLU*

Department of Management Engineering, İstanbul Technical University, Faculty of Management, İstanbul, Türkiye

ARTICLE INFO

ABSTRACT

This study aims to comprehensively explore sustainability practices, guidelines, and emerging trends in the airline industry through an in-depth analysis of their sustainability reviews. A thorough bibliometric analysis of airline sustainability was conducted using the Scopus Database. Additionally, employing text-mining techniques, a meticulous analysis focused on the sustainability reports of the leading 20 airlines compared to the subsequent 20, considered as followers, in the textual examination of sustainability reports. The results revealed thematic disparities between these two cohorts. The top 20 airlines prioritized significant concerns such as safety and other sustainability-related aspects like biodiversity. In contrast, follower companies placed a higher emphasis on financial considerations. This analysis illuminates the primary focuses of airlines and the spectrum of sustainability-related issues. Moreover, offering valuable insights for both researchers and industry practitioners, this study presents a repository of pertinent data related to sustainability practices in the aviation sector. Integrating text mining and bibliometric analysis emphasizes essential facets of airline sustainability, resulting in a comprehensive overview of the research landscape.

Cite this article as: Şapaloğlu İ. Text mining on sustainability reports of top 40 airlines and bibliometric analysis of airline's sustainability. Environ Res Tec 2024;7(X)00–00.

INTRODUCTION

In recent decades, sustainability has emerged as a critical concern for companies worldwide. Organizations now focus on sustainable solutions to utilize resources effectively and contribute to a more habitable world. Solutions prioritizing environmental protection have created new opportunities for gaining a competitive advantage through the efficient use of corporate assets [1].

The aviation industry, in particular, faces significant challenges in balancing operational requirements with environmental stewardship. Airlines have responded to society's increasing demand for environmentally sensitive practices by incorporating sustainability into their operations, regulations, and reporting methods. While the aviation industry offers the opportunity to explore the world, it also places a substantial demand on the world's resources. Passenger traffic, which stood at 34.4 million (domestic and international) in 2003, reached 161 million by November 2022. Over the past decade, the aviation sector has witnessed an energy consumption increase of over 6% [2]. Currently, aviation contributes to more than 2% of the world's greenhouse gas emissions, and in the European Union (EU), this proportion is approximately 3% [3]. Aviation companies have heightened their interest in sustainability to avoid carbon taxes, reduce operational costs, and build a positive reputation. Consequently, they have shifted from merely publishing annual financial reports to including "sustainability re-

---

***Corresponding author.**
*E-mail address: sapaloglu19@itu.edu.tr

2

Environ Res Tec, Vol. 7, Issue. X, pp. 00–00, ????? 2024

ports." Numerous studies in the literature demonstrate the relationship between sustainability reports and corporate reputation [4–8]. Analyzing sustainability reports, a crucial input for corporate reputation is also valuable in examining the current situation and concerns of the aviation industry.

This study comprises two main parts: the first involves researching the subject through both bibliometric analysis and the Scopus database, and the second aims to discuss the findings by applying text mining to the sustainability reports published by the top 40 airline companies. The selection of the top 40 is based on their awareness of the importance of sustainability reporting and their status as flag carrier companies of different countries.

On the regulatory front, governments are pushing companies towards greater environmental friendliness, exemplified by initiatives such as the Emission Trading Scheme (ETS).

This research contributes to the body of knowledge by examining thematic variations between the sustainability reports of the top 20 airlines and those of their competitors. The comparative analysis elucidates the unique priorities of these clusters and underscores thematic patterns in sustainability reporting in the aviation industry. By employing text-mining techniques, the study provides a new perspective, offering a thorough understanding of the linguistic patterns and emphasized topics in sustainability reports. These reports play a crucial role in portraying airline companies as environmentally friendly to their stakeholders. Consequently, this study aims to answer the following research questions:

RQ1: What are the most emphasized topics in the sustainability reports of airline companies?

RQ2: Are there thematic differences between the sustainability reports of the top 20 and the followers?

RQ3: Are there similarities and/or differences in trends between the airline companies' sustainability reports and the literature?

The aviation industry's rapid growth poses challenges in striking a balance between environmental responsibility and operational needs. Grasping the thematic nuances becomes critical as airlines increasingly publish sustainability reports alongside financial reports. The need to understand airline priorities and concerns in the sustainability space is the driving force behind this study. Despite the growing importance of sustainability reporting, there are still few studies systematically comparing the sustainability reports of leading airlines. This research aims to fill this knowledge gap by providing a thorough analysis of thematic differences and commonalities, advancing both academic and practical knowledge in the field.

The study's data comprises information obtained from the Scopus database as well as reports published by various companies. The presentation of the study results will commence alongside the methodology section in the third part.

Literature Review

The civil aviation sector has started to pay attention to its commitment to minimizing and lowering its environmental consequences as one of the major sources of energy use and carbon emissions [9]. Because of this attention, companies, society, governments, and regulators want the aviation industry to be more sustainable. One of the ways aviation companies prove that they are sustainable is through the reports they have published. The practice of non-financial reporting began in the 1970s with social disclosures, then expanded to include broader social and environmental reporting in the 1990s. After the turn of the millennium, these reports increasingly combined and came to be known as sustainability reports [10, 11]. Sustainability reports can be used in impression management to keep the company's credibility high and as a tool for actions being beneficial [12]. One of the methods used in the analysis of sustainability reports is text mining. Text mining is the use of natural language processing and data mining together, it is a way of conducting research with the aim of processing and using all human-specific languages [13].

Seo and Itoh, [14] highlight the potential of text mining in analyzing passenger word-of-mouth to gain insight into the evolution of global airline alliance efforts. Similarly, Tian et al. [15] propose a new methodology that uses text mining and sentiment analysis to evaluate service quality in the airline industry, highlighting the potential to extract valuable insights from social media data. Additionally, Menezes et al. [16] advocate combining quantitative longitudinal data analysis with text mining and qualitative inferences from sustainability reports to improve understanding and decision-making.

In addition to methodological considerations, it is also crucial to understand the specific focus areas in sustainability reports. Moreover, Zhang [17] highlights the importance of comparative studies of sustainability reports in different regions, highlighting the attention paid by European and Asia-Pacific airlines to economic, social, and environmental issues. That study highlights the importance of considering regional differences and specific sustainability issues in the airline industry.

Additionally, the temporal aspect of sustainability reporting is also important. Yang et al. [18] found changing trends in corporate social responsibility reporting among the world's leading airlines, highlighting the dynamic nature of sustainability disclosures over time and across regions. Similarly, Paraschi (2022) [19] discusses the importance of sustainability reporting strategies, especially during crises such as the COVID-19 pandemic, and argues that investments in ESG (Environmental, Social and Governance) and sustainability practices can lead to sustainable profitability and market performance.

In the study of Kim and Kim [20] by analyzing sustainability reports and newspaper articles with text mining, they determined that the concepts of ethical issues, sustainable production, quality, and customer roles are emphasized in the texts analyzed the most. Health and safety, human rights, lowering greenhouse gas emissions, conserving en-
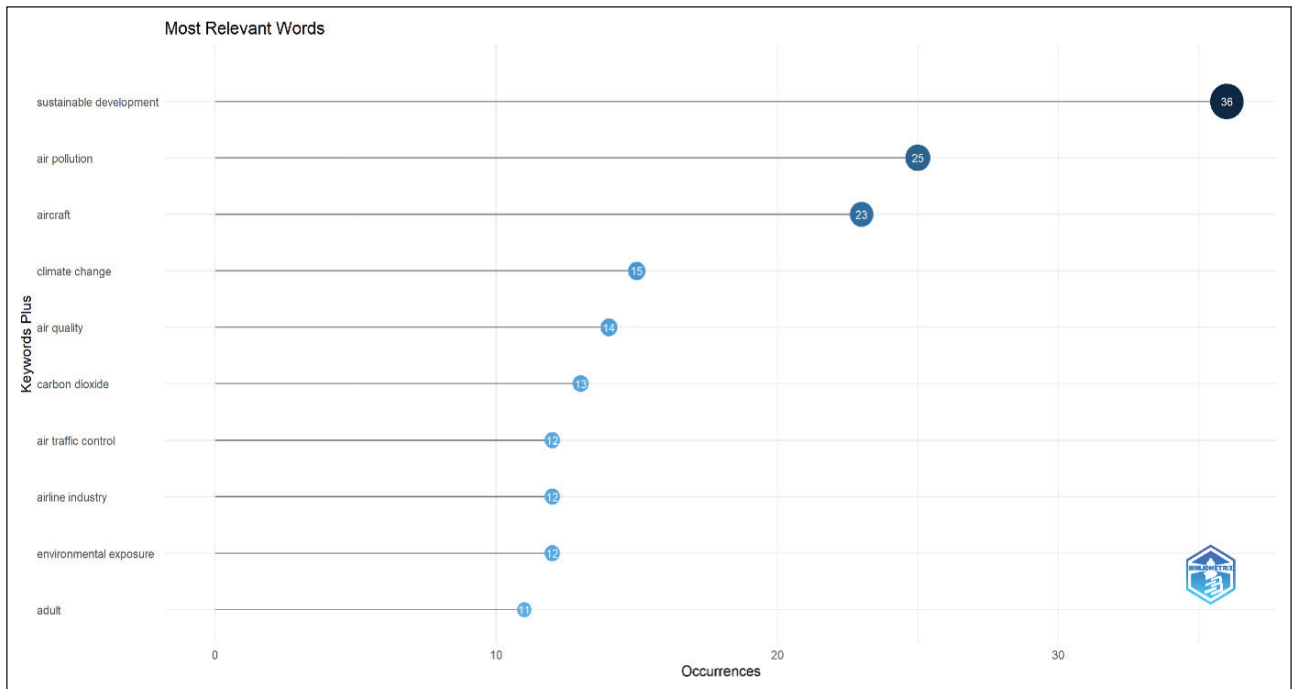
ergy/improving energy efficiency, and community investment are the four sectors that are most heavily focused in the research of Liew et al., [21] that is based on trends on sustainability via text mining. Modapothala and Issac, [22] demonstrated that environmental variables contribute much more to the explanation of the sustainability report based on the text mining results. Zhou et al., (2021) [23] by text mining on company sustainability reports, strive to identify themes and patterns in how container transport businesses disclose sustainability.

Aviation businesses are more likely to participate in such reporting if they are located in nations with solid structures of government and high ethical and ecological standards [24]. The impact of regulators, one of the stakeholders, on sustainability is apparent at this point.

MATERIALS AND METHODS

For the bibliometric analysis stage of the study, the procedures of qualitative content analysis were based on the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flow chart. In the initial phase, 143 papers were identified. After scrutinizing Step 2 and eliminating three duplicate documents, I encountered a partially Chinese document in Step 3. Following this step, I examined the paper abstracts and excluded 12 papers that were either unrelated or loosely connected in Steps 4 and 5.

Text Mining application on the sustainability reports of the top 40 airways followed the below-stated sub-steps:

1. ***Tokenization***: Individual words or terms from the pre-processed text corpus were separated. To create a collection of tokens, the text has to be tokenized, or divided into its component parts [25].

2. ***Stop words removal***: The tokenized corpus was rid of often-used stop words like "*the*," "*is*," and "*and*." It is fair to omit these terms from the analysis because they have little meaning. The stop words phase significantly impacts the findings [26].

3. ***Calculation of term frequency***: The tokenized corpus's term frequency was determined. In addition to providing insights into the major themes and issues featured in the sustainability reports, this stage assisted in identifying the terms that were used the most frequently [27] The most commonly used word groups were determined by looking at the co-occurrence levels of words and the high-frequency words visualized using the frequency histogram and word cloud, as Eles et al. [28] and M.-J. Kim et al. [29] did. Researchers frequently apply frequency normalization because word frequency can strongly depend on document length [30]. One of the widely used term frequency normalization strategies is augmented normalized term frequency (ANTF) [31]. This method was first introduced by Croft [32], who described it as 0.5+0.5* *tf/max* tf, where *tfmax* is the greatest term 3 in a document and tf is the frequency at which a term appears in a text.

Figure 1. Word cloud of 127 documents from scopus.

4. ***Topic modeling***: Latent Dirichlet Allocation (LDA), a topic modeling technique, was used to identify latent subjects in the sustainability reports. Each document received a subject assignment by LDA, which also found keywords related to each topic. The reports' underlying theme structure was made clear by this examination [33]. For each document (*among M*), choose a Dirichlet distribution of K topics (θ^*) based on parameters (α^*). Then, for $n = 1, \ldots, N$, choose:

-A topic zn from a multinomial distribution with parameters (θi).

-A word wn from $p(wn \mid zn, \beta)$, a multinomial distribution conditioned on zn, with parameters (β^*). The joint probability of document w (consisting of N words), the set of N topics z, and the topic mixture θ is [34]:

$$p(w, z, \theta \mid \alpha, \beta) = p(\theta \mid \alpha) \times \Pi[N, n=1] (p(zn \mid \theta) \times p(wn \mid zn, \beta)). \quad (3.1)$$

I utilized R's "tm" package to facilitate document transfer and corpus construction. Widely acknowledged as a proficient text mining tool, the "tm" package excels in handling corpora, preprocessing data, creating term-document matrices, and more [35]. The textual data underwent a meticulous cleansing process to yield relevant results. This involved the removal of punctuation marks, numerals, and excessive spaces between pages and paragraphs. Additionally, all letters were converted to lowercase for uniformity. To further refine the dataset, stemming was applied to exclude terms with similar meanings, transforming them into their base forms. Lastly, common English stop words such as '*the*' and '*a*' were excluded from the analysis.

RESULTS

Bibliometric Analysis

As explained in the Methodology section, the study consists of bibliometric analysis and analysis of sustainability reports. In the first stage of the analysis, 143 documents, languages other than English, duplications, and those other than conference papers, articles, and book chapters were removed from the data set. As a result, 127 documents were used for analysis.

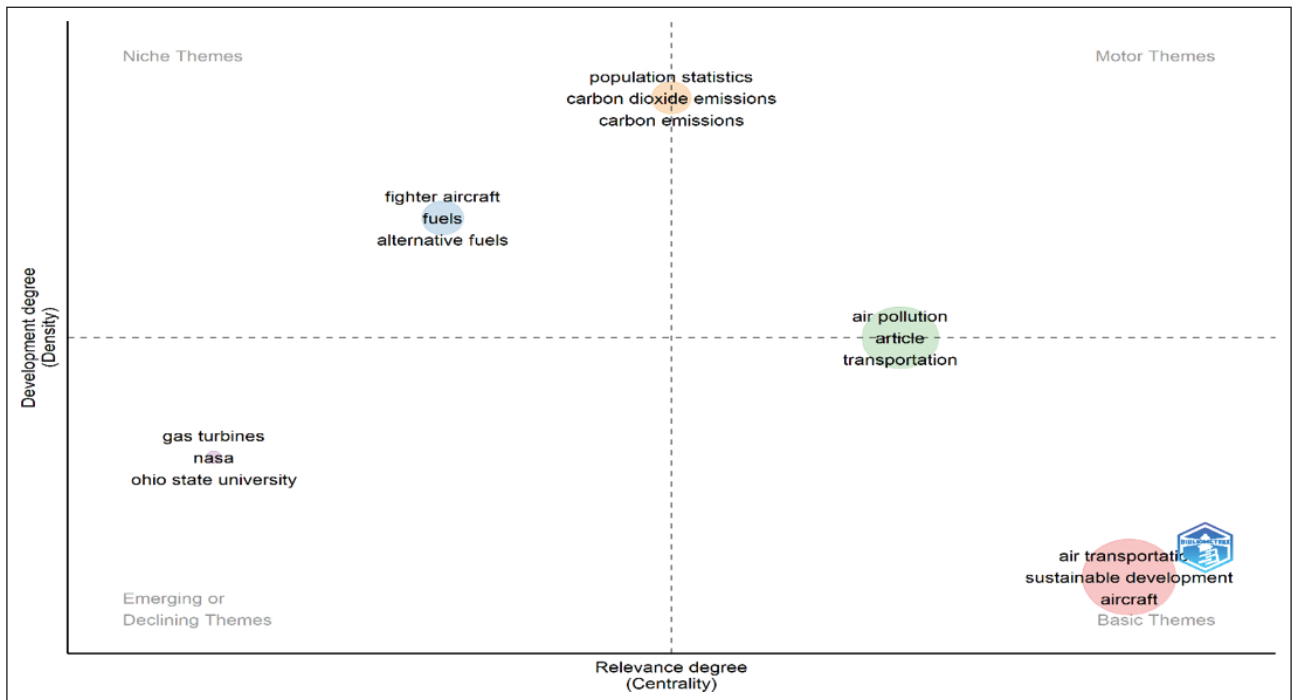**Figure 2**. Most relevant words.

Figure 3. Thematic clustering of documents from the literature.

As seen in Figure 1, "environmental sustainability" seems to be more prominent in the studies in the context of sustainability related to the aviation industry. On the other hand, it is observed that economic sustainability and human health and quality concerns are also emphasized.

According to the most relevant words in Figure 2, the word meaning is sustainable development. This especially shows the importance of SDGs even in terms of airways. There are positive reflections on the performance of the companies that act with the awareness of the SDGs [36, 37].

As has been pointed out in Figure 3 the thematic map had five clusters. When the results were considered, niche themes were fighter aircraft and alternative fuels. Emerging themes were gas turbines and NASA. The Basic (developing) themes were air transportation, sustainable development, and aircraft. Motor (developed) themes although the boundaries are not exactly clear were air pollution, pollution, carbon emission, and carbon dioxide emissions.

As depicted in Table 1, illustrating the frequencies and country-specific distribution of documents within the top

Table 1. The list of the top 40 airways by country

Airways	Country	Airways	Country
Id1	Qatar	Id21	Finland
Id2	Singapore	Id22	Bahrain
Id3	UAE	Id23	Thailand
Id4	Japan	Id24	USA
Id5	Australia	Id25	Spain
Id6	Japan	Id26	Ethiopia
Id7	Türkiye	Id27	Thailand
Id8	France	Id28	Greece
Id9	Korea	Id29	Singapore
Id10	Switzerland	Id30	New Zealand
Id11	Britain	Id31	Indonesia
Id12	UAE	Id32	Australia
Id13	China	Id33	Korea
Id14	China	Id34	Saudi Arabia
Id15	Germany	Id35	USA
Id16	China	Id36	Fiji
Id17	Netherlands/France	Id37	Oman
Id18	Taiwan	Id38	Kazakhstan
Id19	USA	Id39	Canada
Id20	India	Id40	Irland

Source: [38].

Table 2. Academic documents frequency by country

Academic documents on Scopus	
Country	**Frequency**
Usa	48
China	21
Italy	17
Germany	15
Netherlands	14
Spain	14
Malaysia	10
India	9
UK	9
France	8

Source: Scopus database.

Table 3. Variables of sustainability reports analysis

Variable	Definition
Txt files	An independent document id spanning from 1 to 40
Text	Text of the under-inquiry work
Cluster 1	Data of top 20 companies [1–20]
Cluster 2	Data of followers companies [21–40]

40 airline lists, it is noteworthy that China, the Netherlands, and the USA emerge as prominent contributors, as evident in Table 2, where a substantial number of academic studies have been conducted. Additionally, the airlines from these countries hold positions within the Top 20 airline list. Top 20 airlines list gathered from Skytrax website [38].

The next part of the study aims to answer the research questions by analyzing the companies' sustainability reports.

Text Mining on Sustainability Reports
The analysis at this stage compares the top 20 companies and the other 20 companies that are followers of them. To carry out the analysis in this section, it is necessary to prepare the data first. In this context, the sustainability reports in the dataset, which are pdf, have been converted to text file format. The average word count of the 40 files examined in this section is 7302.154.

Companies in the top 40 in the research were divided into two clusters: top 20 and followers. In this way, it is aimed to reveal the thematic differences behind the success of the top 20. It is aimed to compare these two clusters and to provide the opportunity to see a shortcoming of the following companies. The varibles used in the analysis process are shown in Table 3. In the text files, many elements could not be analyzed because these files were company reports. To get rid of all these, the cleaning processes mentioned in the 3rd part were carried out.

Findings of Analysis of the Sustainability Reports
The analysis findings of the documents listed as 40 separate document IDs and divided into two clusters, the first 20 and the next 20, will be discussed in this section.

When Figure 4 is examined, it is possible to have an idea about the priorities of the top 20 companies. For example, while safety is the most important issue for the top 20 companies, on the other hand, finances have emerged for the followers. It is not surprising that a concept like biodiversity is in the top 20 and not in the followers. This may show that these companies are aware that they cannot be sustainable by using only resources less.

When the bi-gram words in Table 4 are examined, it is seen that there ara distinctive differences between the top 20 and the followers in parallel with Figure 4. While the top 20 focus mainly on sustainability, the followers focus more on financial issues.

Drawing insights from the content and frequency of the text data, themes outlined in Table 5 came to the forefront. A few overarching observations emerge from this analysis:

Topic 1 elucidates the intricate interplay among environmental responsibility, ethical conduct, security considerations, and the realm of business operations. The relationship between sustainability and employees, as well as monetary concerns, is highlighted in Topic 2. The significance of data and reporting on air, carbon, and the environment is demonstrated in Topic 3. Topic 4 demonstrates how companies integrate their sustainability objectives with corporate responsibili-

6

Environ Res Tec, Vol. 7, Issue. X, pp. 00–00, ????? 2024

Figure 4. Word/frequency graphs of the top 20 and next 20 companies' reports (left side top 20, right side followers).

Table 4. Bi-gram word frequencies of the top 20 and followers

Top 20		Followers	
Safety standards	455	Financial statement	470
Climate change	349	Financial assets	390
Sustainable development	319	Aviation fuel	349
Carbon emission	312	Jet fuel	346
Fuel efficiency	310	Health safety	346
Waste management	306	Supply chain	341
Corporate governance	279	Fuel efficiency	339
Employee engagement	262	Carbon emission	320
Risk management	237	Cash flow	220
Environmental impact	192	Corporate governance	217

Table 5. Topic modeling analysis on airline sustainability reports (id1-id40)

Topic 1: statement, safety, operation, sustain, green

Topic 2: financial, flight, sustainable, operation, employees

Topic 3: report, air, carbon, green, environment

Topic 4: operation, sustainable, board, social, efficiency

Topic 5: consolidation, operation, governance, safety, sustainable

Figure 5. Word cloud of the sustainability reports [id1-id40].

no term related to the employees or other social environment elements. In this context, in Figure 5, words such as human, community, employee, and training in the word cloud are prominent.

ty and social responsibility. The effects of consolidation are highlighted in Topic 5, along with the relationship between business performance, security, and sustainability.

When evaluated together with the literature findings that were mentioned in bibliometric analysis, it is seen that companies do not only refer to environmental and economic sustainability but also to social sustainability to emphasize that they are aware of their responsibilities towards society and their stakeholders.

Figure 2, which shows the most relevant words in the literature, is mentioned in the reports, although there is

CONCLUSION

This study employs both bibliometric analyses from the literature and text-mining techniques on sustainability reports from the top 40 airlines, providing a comprehensive examination. The robustness of the findings is strengthened by this dual approach, offering a thorough understanding of thematic differences. The integration of quantitative and qualitative methodologies contributes to the expansion and depth of knowledge within the field of airline sustainability.

Results obtained from the bibliometric analysis reveal a pronounced emphasis on environmental aspects in sustainability studies within the aviation industry. Countries such as China, the Netherlands, India, and the USA emerge as key contributors to the academic discourse, featuring a

multitude of scientific inquiries, with their respective airlines prominently represented in the cohort of the top 40 airlines. A meticulous nation-wise assessment of scientific document frequency underscores these trends.

The research findings hold significant implications for practitioners and policymakers in the aviation industry. Airlines can refine their strategies to align with evolving sustainability goals by thoroughly understanding the thematic priorities highlighted in sustainability reports. Using these insights, managers can enhance their reporting procedures and ensure that sustainability projects are communicated to stakeholders more comprehensively and effectively.

Additional insights were derived from text-mining analysis of the sustainability reports of the top 20 airlines and their followers. The analysis reveals distinct themes in the reports of the top 20 airlines and those of the followers. Safety, environmental sustainability, and social responsibility take precedence in the reports of the top 20 airlines, while followers place greater emphasis on financial considerations. The reports of the top 20 airlines also feature terms like biodiversity more frequently, demonstrating their recognition of the importance of considering a broader range of sustainability issues. As discussed by Paraschi (2022) [19], sustainability reporting strategies and investments in ESG and sustainability practices can lead to sustainable profitability and market performance. Conversely, Yang et al. (2020) [18] state in their study that there are changing trends in sustainability reports or corporate social responsibility reports, emphasizing the need for airline companies to adapt to maintain competitiveness. This study did not identify clear distinctions between the top 20 airlines and their followers regarding regional differences, as both clusters include airlines from countries with both Western and Eastern cultures.

While this study provides valuable insights, it is essential to acknowledge its limitations. Firstly, the analysis is based on sustainability reports made available to the public, which may or may not be comprehensive. Additionally, the study's focus on the top 40 airlines might overlook the diversity present in smaller or regional airlines. The quality and format of the reports may impact the efficacy of text-mining techniques, influencing the extent of the analysis. Furthermore, the study does not delve into regional differences in sustainability priorities, a potential avenue for further investigation.

Future studies may explore whether regional differences exist in reports based on the countries or regions where airline companies are located. Additionally, examining whether company structures or management types affect differentiation in reports could be a valuable area for investigation.

DATA AVAILABILITY STATEMENT

The author confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

USE OF AI FOR WRITING ASSISTANCE

Not declared.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1]　R. Wüstenhagen and M. Starik, "Sustainable Innovation and Entrepreneurship Reducing permitting risks of large wind energy projects in Switzerland View project UNISG former research work View project," Edward Elgar Publishing Limited, 2008. [CrossRef]

[2]　A. Hadi-Vencheh, P. Wanke, A. Jamshidi, and Z. Chen, "Sustainability of Chinese airlines: A modified slack-based measure model for CO2 emissions," in Expert Systems, Blackwell Publishing Ltd, 2020.

[3]　M. Efthymiou, and A. Papatheodorou, "EU Emissions Trading Scheme in aviation: Policy analysis and suggestions," Journal of Cleaner Production, Vol. 237, Article 117734, 2019. [CrossRef]

[4]　D. L. Brown, R. P. Guidry, and D. M. Patten, "Sustainability reporting and perceptions of corporate reputation: An analysis using fortune," in Sustainability, environmental performance and disclosures, Emerald Group Publishing Limited, 2009. [CrossRef]

[5]　Zimon, A. Arianpoor, and M. Salehi, "Sustainability reporting and corporate reputation: the moderating effect of CEO opportunistic behavior," Sustainability, Vol. 14(3), Article 1257, 2022. [CrossRef]

[6]　Y. A. Abbas, W. Mehmood, Y. Y. Lazim, and A. Aman-Ullah, "Sustainability reporting and corporate reputation of Malaysian IPO companies," Environmental Science and Pollution Research, Vol. 29(52), pp. 78726–78738, 2022. [CrossRef]

[7]　M. D. Odriozola, and E. Baraibar-Diez, "Is corporate reputation associated with quality of CSR reporting? Evidence from Spain," Corporate Social Responsibility and Environmental Management, Vol. 24(2), pp. 121–132, 2017. [CrossRef]

[8]　D. L. Brown, R. P. Guidry, and D. M. Patten, "Sustainability reporting and perceptions of corporate reputation: An analysis using fortune," in Sustainability, environmental performance and disclosures, Emerald Group Publishing Limited, 2009. [CrossRef]

[9]　A. Hadi-Vencheh, P. Wanke, A. Jamshidi, and Z. Chen, "Sustainability of Chinese airlines: A modified slack-based measure model for CO2 emissions," Expert Systems, Vol. 37(3), Article e12302, 2020. [CrossRef]

[10]　M. S. Fifka, "Corporate Responsibility Reporting and its Determinants in Comparative Perspective – a Review of the Empirical Literature and a Meta-analysis," Business Strategy and the Environment,

Vol. 22(1), pp. 1–35, 2013. [CrossRef]

[11]   M. Zieba, and E. Johansson, "Sustainability reporting in the airline industry: Current literature and future research avenues," Transportation Research Part D: Transport and Environment, Vol. 102, Article 103133, 2022. [CrossRef]

[12]   P. Bansal, and G. Kistruck, "Seeing is (not) believing: Managing the impressions of the firm's commitment to the natural environment," Journal of Business Ethics, Vol. 67, pp. 165–180, 2006.

[13]   T. Kaşikçi, and H. Gökçen, "Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi," Bilişim Teknolojileri Dergisi, Vol. 7(1), pp. 25–32, 2014.

[14]   G.-H. Seo, and M. Itoh, "Perceptions of customers as sustained competitive advantages of global marketing airline alliances: A hybrid text mining approach," Sustainability, Vol. 12(15), Article 6258, 2020.

[15]   X. Tian, W. He, C. Tang, L. Li, H. Xu, and D. Selover, "A new approach of social media analytics to predict service quality: evidence from the airline industry," Journal of Enterprise Information Management, Vol. 33(1), pp. 51–70, 2020. [CrossRef]

[16]   L. M. de Menezes, A. B. Escrig-Tena, and J. C. Bou-Llusar, "Sustainability and Quality Management: has EFQM fostered a Sustainability Orientation that delivers to stakeholders?," International Journal of Operations & Production Management, Vol. 42(13), pp. 155–184, 2022. [CrossRef]

[17]   X. Zhang, "Communicating social responsibilities through CSR reports: Comparative study of top European and Asia-Pacific airlines," PLoS One, Vol. 16(10), Article e0258687, 2021. [CrossRef]

[18]   L. Yang, C. S. B. Ngai, and W. Lu, "Changing trends of corporate social responsibility reporting in the world-leading airlines," PLoS One, Vol. 15(6), pp. e0234258, 2020. [CrossRef]

[19]   E. P. Paraschi, "Why ESG Reporting is Particularly Important for the Airlines during the Covid-19 Pandemic," Journal of Business and Management Studies, Vol. 4(3), pp. 63–67, 2022. [CrossRef]

[20]   D. Kim, and S. Kim, "Sustainable supply chain based on news articles and sustainability reports: Text mining with Leximancer and DICTION," Sustainability (Switzerland), Vol. 9(6), Article 1008, 2017. [CrossRef]

[21]   W. Te Liew, A. Adhitya, and R. Srinivasan, "Sustainability trends in the process industries: A text mining-based analysis," Computers in Industry, Vol. 65(3), pp. 393–400, 2014. [CrossRef]

[22]   J. R. Modapothala, and B. Issac, "Study of economic, environmental and social factors in sustainability reports using text mining and Bayesian analysis," in 2009 IEEE Symposium on Industrial Electronics & Applications, pp. 209–214, 2009. [CrossRef]

[23]   Y. Zhou, X. Wang, and K. F. Yuen, "Sustainability disclosure for container shipping: A text-mining approach," Transport Policy (Oxford), Vol. 110, pp. 465–477, 2021.

[24]   M. Kılıç, A. Uyar, and A. S. Karaman, "What impacts sustainability reporting in the global aviation industry? An institutional perspective," Transport Policy (Oxford), Vol. 79, pp. 54–65, 2019. [CrossRef]

[25]   S. Vijayarani, and R. Janani, "Text mining: open source tokenization tools-an analysis," Advanced Computational Intelligence: An International Journal, Vol. 3(1), pp. 37–47, 2016. [CrossRef]

[26]   D. Munková, M. Munk, and M. Vozár, "Data pre-processing evaluation for text mining: transaction/sequence model," Procedia Computer Science, Vol. 18, pp. 1198–1207, 2013. [CrossRef]

[27]   A. Correia, M. F. Teodoro, and V. Lobo, "Statistical methods for word association in text mining," Recent Studies on Risk Analysis and Statistical Modeling, Springer, pp. 375–384, 2018. [CrossRef]

[28]   P. T. Eles, B. Pennell, and M. Richter, "Assessing NATO policy alignment through text analysis: An initial study," in 2016 International Conference on Military Communications and Information Systems (ICMCIS), IEEE, 2016, pp. 1–7. [CrossRef]

[29]   M.-J. Kim, K. Ohk, and C.-S. Moon, "Trend analysis by using text mining of journal articles regarding consumer policy," New Physics: Sae Mulli, Vol. 67(5), pp. 555–561, 2017. [CrossRef]

[30]   V. Lertnattee and T. Theeramunkong, "Effect of term distributions on centroid-based text categorization," Information Science (NY), Vol. 158, pp. 89–115, 2004. [CrossRef]

[31]   J. Gadge, S. Sane, and H. B. Kekre, "N-layer approach to web information retrieval," International Journal of Applied Information Systems, Vol. 5, pp. 45–49, 2013. [CrossRef]

[32]   W. B. Croft, "Document Retrieval system," Information Technology, Vol. 2, pp. 1–21, 1983.

[33]   E. S. Negara, D. Triadi, and R. Andryani, "Topic modeling twitter data with latent Dirichlet allocation method," in 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), IEEE, 2019, pp. 386–390. [CrossRef]

[34]   Y. Haralambous, "Text mining methods applied to mathematical texts," 2012. Accessed on Jun 11, 2023. https://hal.science/hal-01864536

[35]   I. Feinerer, K. Hornik, and D. Meyer, "Text mining infrastructure in R," The Journal of Statistical Software, Vol. 25(5), pp. 1–54, 2008. [CrossRef]

[36]   P. A. Khan, S. K. Johl, and S. Akhtar, "Firm sustainable development goals and firm financial performance through the lens of green innovation practices and reporting: a proactive approach," Journal of Risk and Financial Management, Vol. 14(12), Article 605, 2021. [CrossRef]

[37]   D. L. Ramos, S. Chen, A. Rabeeu, and A. B. Abdul Rahim, "Does SDG Coverage Influence Firm Performance?," Sustainability, Vol. 14(9), Article 4870, 2022. [CrossRef]

[38]   World Airline Awards, "World's Top 100 Airlines 2022 | SKYTRAX." Accessed on Jun 11, 2023. https://www.worldairlineawards.com/worlds-top-100-airlines-2022/