

A Bibliometric of Sentiment Analysis in Tourism Industry during COVID-19 Pandemic

Asra Yaqub

Student at the Computer Science Department
College of Engineering, Effat University
Jeddah, Saudi Arabia
asyaqub@effat.edu.sa

Huda Thalib

Student at the Computer Science Department
College of Engineering, Effat University
Jeddah, Saudi Arabia
huthalib@effat.edu.sa

Tayeb Brahimi

Associate Professor at the College of Engineering
Energy and Technology Research Centre
Natural Science, Natural Science, Mathematics and Technology Unit
Effat University, Jeddah, Saudi Arabia
tbrahimi@effatuniversity.edu.sa

Akila Sarirete

Assistant Professor at the Computer Science Department
College of Engineering, Effat University
Jeddah, Saudi Arabia
asarirete@effatuniversity.edu.sa

Abstract

The purpose of this paper is to explore, visualize, and analyze the state-of-the-art research of sentiment analysis in the tourism industry during COVID-19 using bibliometrics. The information source is obtained from the Scopus database from January 1, 2020, to November 27, 2021. A sample of 1690 documents was obtained within this period. Topic areas with titles, keywords, and abstract criteria in sentiment analysis and travel and tourism were used as a reference for extracting results. There was a significant increase in papers published in 2020 and 2021 related to COVID-19 and sentiment analysis. The top five most active countries were the United States (310 documents and 1799 citations), followed by the United Kingdom (170 documents and 12674 citations), China (240 documents and 1297 citations), Australia (120 documents and 430 citations), and Germany (44 documents and 172 citations). The maximum number of occurrences was "social media (401)" followed by "sentiment analysis (149)", "tourism (116)", "Covid-19 (131)", and "Twitter (77)". This study provides an overview of sentiment analysis and shows the trends in the research and publication with the tourism industry.

Keywords

Sentiment Analysis, Tourism, COVID-19, Bibliometric, Text Mining

1. Introduction

The advent of Web 2.0 has led to an increase in the generation of humongous amounts of data as more people resort to connecting and sharing opinions on social media. Social Media has become a hub for generating sentiment-rich data in the form of tweets, status updates, blog posts, comments, reviews, etc (Kharde 2016). Various computational tools exist to refine the data to monitor reputation and understand consumer trends (Mohamed Hussein 2018). Social media has changed the way people share information, and its application in sentiment analysis has proved beneficial for businesses (Sheth 2020). Tourism nowadays is viewed from many perspectives, such as in social, psychological, cultural, and economic contexts (Stylidis and Terzidou 2013) (Roman et al. 2020)

The COVID-19 pandemic has had a catastrophic effect on the economic sectors of each country, and this effect was high on the tourism industry as air, sea, and land connections were closed for leisure travel (Polyzos et al. 2020). Tourists rely on social media applications to learn about destinations' health and guidelines, making travel one of the most discussed topics on social media (Neidhardt et al., 2017; Travel mail Reporter, 2013). Businesses rely on customer feedback and experiences to determine their growth, reputation, and industry trends.

The use of bibliometric analysis, data visualization, and mapping to study statistical behavior and trends based on visualization tools and databases like Web of Science (WoS) and Scopus has increased since the pandemic's emergence (Bankar and Lihitkar 2019) (Hu et al. 2020). The paper aims to explore research on the role of sentiment analysis in the field of tourism industry during COVID-19 pandemic by reviewing existing knowledge and research that might assist in identifying gaps and future directions for the industry.

The remainder of the present paper is structured as follows: section two examines relevant literature review on sentiment analysis and bibliometrics; section three outlines the research methodology and presents the method used in this study; section four outlines the findings and discussion, and section five highlights the major conclusions of the study

2. Literature Review

Social media platforms have eased communications amongst communities allowing them to disseminate information easily (Kim Sin and Yoo-Lee 2014). Sentiment analysis contributes to understanding human emotions via social media applications, and people's behaviors can be seen from their engagement on these platforms (Ji et al. 2016). These media applications have been employed in various application domains, including tourism (Ainin et al. 2020) and health (Rodrigues et al. 2016), to analyze highly dynamic and real-time data trends (Chaudhary and Naaz 2017).

Tourism as an industry is based on customer experience for its growth and reputation. The constant use of social media allows users to generate big data, which can be a source to understand the customer's needs, feelings, and desires. Emotions play a vital role in the tourist experience (Mcintosh and Siggs 2005). Due to the rapid spread of opinionated posts on all social media networks such as Twitter and Facebook has become an important field for marketers to research. There are two different types of information available online. One of them is user-generated content, including opinions, images, or videos uploaded by people. The second type is the interaction between the users, including organizations, products, etc (Flores et al. 2021). Research showed that during the beginning of the pandemic, tourism-related tweets were the most active from 16-18 March 2020. (Obembe et al. 2021) Analyzing such large volumes of valuable subjective information would be highly beneficial in decision making as it analyzes the polarities of sentiments for the extracted data from social media. The internet has significantly transformed people's communication level as it has become a crucial platform for people to express their emotions and opinions. In addition to Instagram, Google Reviews is also often used by travelers to reference more detailed tourism destination information, such as reviews of certain hotels (Phillips et al. 2017). Hospitality and online reviews are becoming a trend and a guide for tourism (Schuckert et al. 2015). The importance of reviews on social media has had a huge impact on consumers as they attain information with just a single click. Moreover, the data from TripAdvisor has indicated that 77% of people check the former guest's reviews before confirming their reservation in a hotel. (Flores et al. 2021)

To a great extent, tourism depends on the range and type of hotels available at the destination. Hotels are a core area of the tourism industry and play a distinctive role in developing this ever-expanding industry (Thomas and Devi 2017). The tourism industry and its associated fields experienced a decline during this period and were the frontline of Covid-19 impact. (Demola et al. 2021). It is estimated that the revenue loss of airlines globally due to Covid-19 in 2020 was

\$370 billion (Mazareanu 2021). The pandemic period raised questions about the sustainability of tourism sectors, such as cruises and aviation (Cretu et al. 2021)

Sentiment Analysis or SA is one of the most efficient techniques as it contributes to analyzing and comprehending people's emotions by evaluating their engagement. Since people express their opinions freely, this technique will provide accurate real-time trends (Alamoodi et al. 2021). Research showed that the World Health Organization (WHO) failed to provide public guidance, and these results were perceived through Sentiment Analysis. (Chandra and Krishna 2021).

Lu and Zheng (2020) conducted a study examining the public sentiments towards cruise tourism during the beginning of 2020, and it also aimed at understanding the impact of Covid-19 on this sector. The results stated that the people were frustrated with quarantine and isolation, making them more eager to travel. The author also observed that such restrictions could lead people to develop an interest in niche cruises and start distancing from mass cruises.

The pandemic severely affected the travel and tourism industry, but it has been significantly underestimated as there were very few studies conducted on this concern. (Škare et al. 2021). The historical effects of past epidemics and pandemics have assisted in recognizing the parameters of COVID-19. Previously, after World War II, several virus outbreaks have affected the worldwide tourism industry. SARS and H1N1 epidemics are discussed, focusing on the tourism demand in various countries such as Europe, South Africa, the US, etc. The author of this research claims that this is the first attempt that measures the impact of the Covid 19 outbreak on tourism worldwide. Therefore, the estimation of this impact is measured using the data of previous coronavirus outbreaks, including SARS, H1N1, Ebola, and some of the most disastrous epidemic outbreaks. It is also highlighted that such pandemic outbreaks are much more severe than what people anticipate because the destructive impact caused by epidemics is nowhere close. (Škare et al. 2021)

Figure 1 shows the primary themes discussed in this paper, such as management, methodology, tourism sectors, and relevant data. Customer experiences are crucial for the tourism industry as it promotes their growth and reputation. This creates a requirement for the management to understand the customer's needs, feelings, and desires to give better services and facilities. Hence, SA has emerged as a prominent methodology for tourism research. It is essential to highlight that several researchers have used SA to understand and predict the future in the tourism industry context. (Manosso and Ruiz 2021) Furthermore, some tourism researchers have divided the sentiments of tourists into two broad categories: positive and negative. Further, each category is classified into various basic emotions such as regret, anger, fear, joy, sadness, etc. (Mitas et al., 2012).

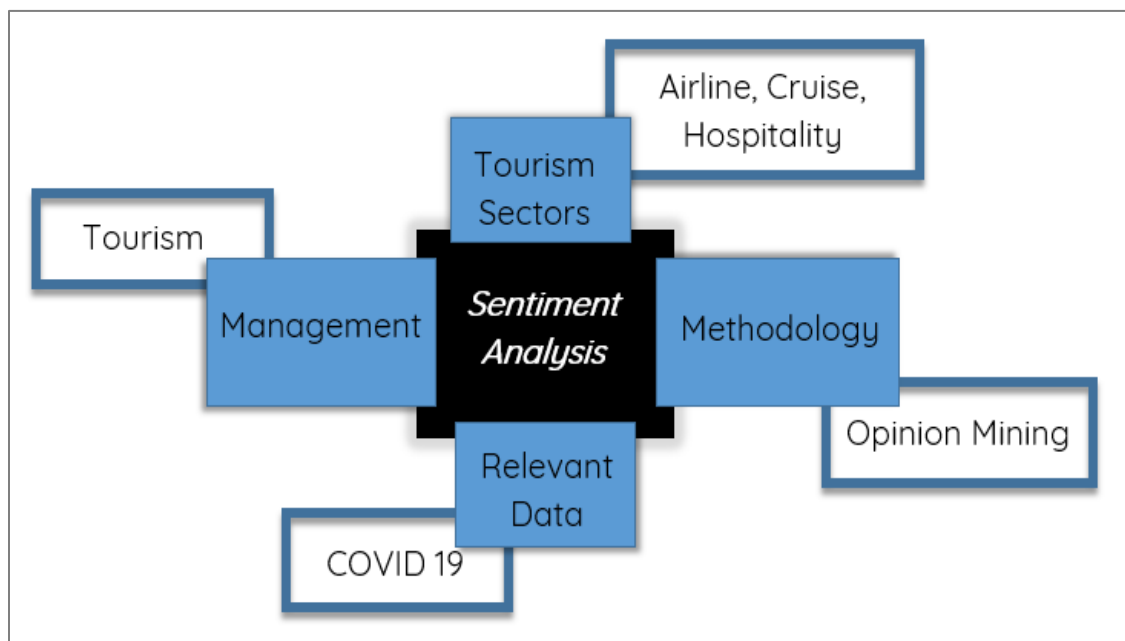


Figure 1. Themes discussed under Sentiment Analysis

3. Methodology

For our sources, we relied on Scopus, a large multidisciplinary comprehensive database consisting of peer-reviewed literature rich in scientific and academic information (Klapaka and Slaby 2018), to extract articles relevant to sentiment analysis in the tourism industry during covid-19. A search query with the following words was used: TITLE-ABS-KEY (("Sentiment Analysis" OR "Text Mining" OR "Opinion Mining" OR "Twitter" OR "Social Media") AND ("Sentiment Analysis" OR "Text Mining" OR "Opinion Mining" OR "Twitter" OR "Social Media") AND ("Tourism*" OR "Tourist" OR "Travel*" OR "Touristic" OR "Excursionist" OR "pilgrim*")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR, 2020)) AND (LIMIT-TO (LANGUAGE , "English")). 24134 document results were retrieved by Scopus when all fields were included. However, we limited the output to a combined field that searches abstracts, keywords, and document titles, which retrieved 1690 documents between January 1, 2020 to November 27, 2021. In the Web of Science (WoS) and using the same query 1809 documents were found, however the number of selected records in WoS was limited to 500 only while the limit for Scopus was 2000. As a result, all the present paper's analysis will only be based on the Scopus database. Using the document retrieved from the Scopus database, which was exported in CSV format, the open-source application VOSviewer (version 1.6.17) (Eck and Waltman, 2010) was used to gain a more in-depth understanding of sentiment analysis in the tourism industry to analyze current research trends. VOSviewer uses the co-occurrence matrix to build a map and generate a similarity matrix. The similarity matrix is then used to generate a VOSviewer map. Finally, the map is translated, rotated, and mirrored in the third phase to make it more realistic. Figure 2 depicts the major stages of the bibliometric analysis conducted for this study.

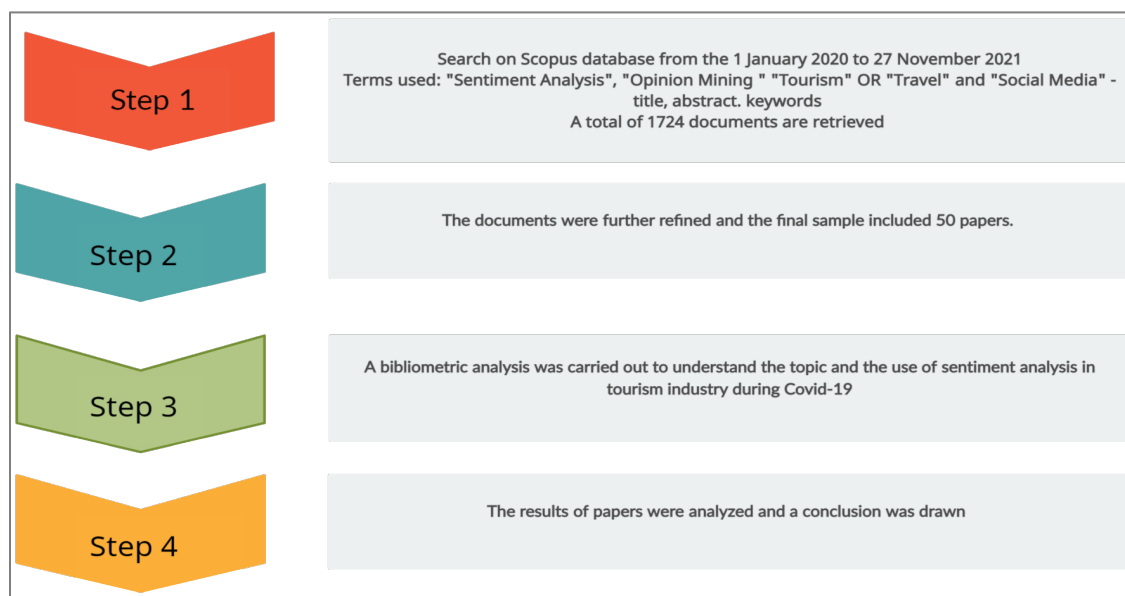


Figure 2. Main stages of the Bibliometric analysis

4. Results and Discussion

Figure 3 shows the growing attention to the sentiment analysis field from 2007 to 2021. At the beginning of the period, there were very few papers published. Still, after 2011 the field saw an increase in research, becoming one of the most active research areas in natural language processing (NLP), data, and web mining. While the role of sentiment analysis is remarkable in psychology research, it has several applications, such as understanding customer behavior, clinical medicine, and building better prediction models for trading stock (Chandra and Krishna 2021).

Every source covered by the Scopus is assigned to one subject category, and there are 11 different subject categories related to our investigated documents. Out of 11, figure 4, four subject categories contain more than 60% of documents. Most publications were in the 'Social Science' category (~23.2%) followed by 'Businesses and Management' (~18%). Computer sciences (~11.7%), Environmental Sciences (~8.4%), and Engineering (~7.2%). Other fields such as Medicine, Energy, Decision Science, Arts and Humanities, and Earth and Planetary Science had publications, but they were low. Although sentiment analysis is a subfield of computer science, it has spread to

management and social sciences in recent years because of its importance to business and society (Liu 2015).

There has been an increase in publications during Covid-19, reaching 892 published documents in 2021. Computer technologies have provided authorities with opportunities to fight outbreaks and have a remarkable role, especially in sentiment analysis for social media (Singh et al. 2018). The importance is mainly due to their tremendous role in analyzing public sentiments. Various research have indicated that outbreaks and pandemics would have been promptly controlled if experts considered social media data (Singh et al. 2018). COVID-19 remains a controversial global topic in social media (Pastor 2020). Therefore, sentiment analysis in studying pandemics, such as COVID-19, is important based on recent events.

Figure 5 shows the top 10 most active countries publishing in this area. The United States has dominated the chart with more than 313 documents, followed by China (245 documents) and the United Kingdom (171 documents). Australia, India, and Spain have similar numbers of publications. Developed countries, especially the USA and the United Kingdom, appear to be more proactive in global food and food-related research topics (Chen et al. 2017, Vanga et al. 2015; Monasterolo et al. 2016; Tan et al. 2014).

Figure 6 shows that Journal articles contributed to a maximum number of publications for articles (72%) followed by conference papers that constituted (14.4%). Other document types included reviews (3.8%), conference reviews (2.9%), book chapters (2.5%), letter (1.3%), editorial (0.9%), note (0.8%), erratum (0.3%), book (0.2%) and others (0.3%).

Figure 7 illustrates the keywords cloud. It summarizes the research aim visually, and the size of each keyword indicates its relative use.

With the help of VOSviewer, an interactive map based on Scopus articles, authors, research institutes, and countries has been constructed. The network analysis with author keywords is depicted in Figure 8. The circles represent the nodes in the network. An item's label will be larger if it is considered important. The color of an object is determined by the cluster to which it belongs, and the lines represent the connections between objects. Objects are colored according to their clusters. 179 out of the 4703 keywords with a minimum occurrence of 5 met the threshold, while the remaining papers did not. The maximum number of occurrences was "social media (401)" followed by "sentiment analysis (149)", "tourism (116)", "Covid-19 (131)", and "Twitter (77)". Figure 7 also shows 12 clusters grouped by color.

Figure 9 shows the network analysis of country co-authorship. Out of 135 countries, 58 meet the threshold of a minimum of 5 documents by country. The output shows the top five most active countries starting with the United States (313 documents and 1799 citations), followed by the United Kingdom (170 documents and 12674 citations), China (240 documents and 1297 citations), Australia (120 documents and 430 citations), and Germany (44 documents and 172 citations).

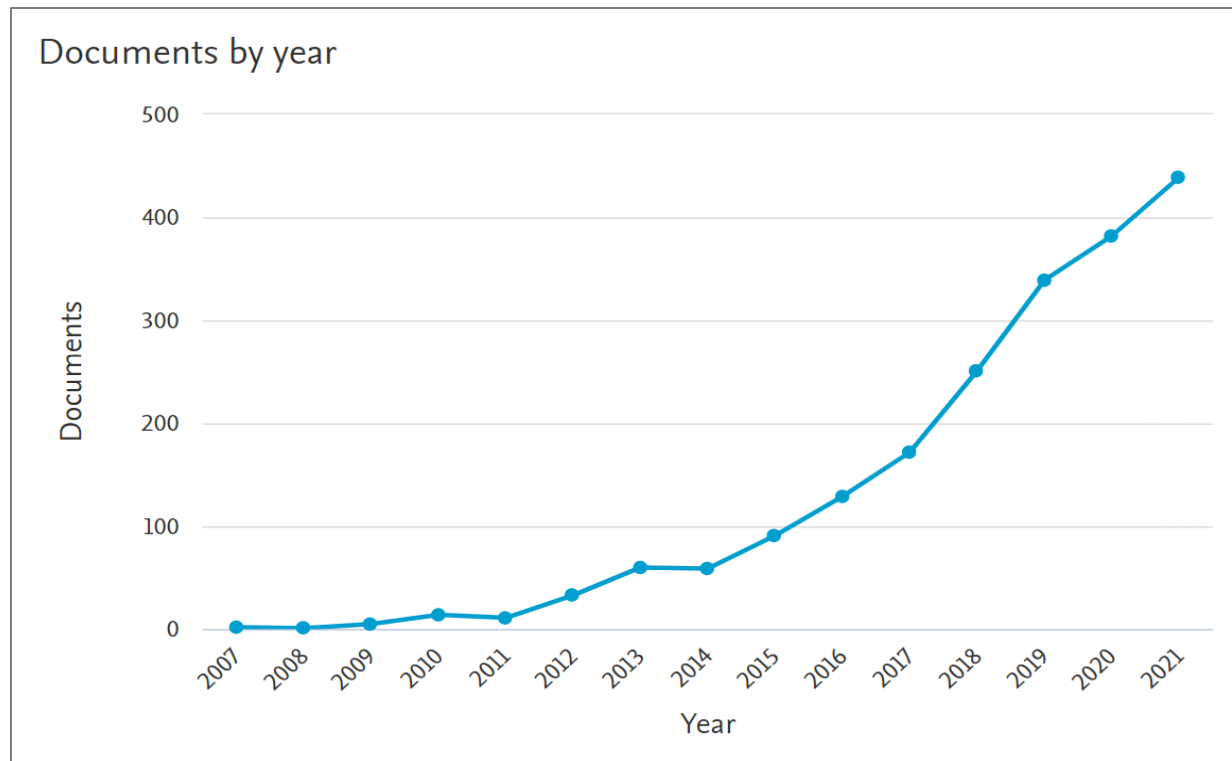


Figure 3. Distribution of sentiment analysis documents published over the years

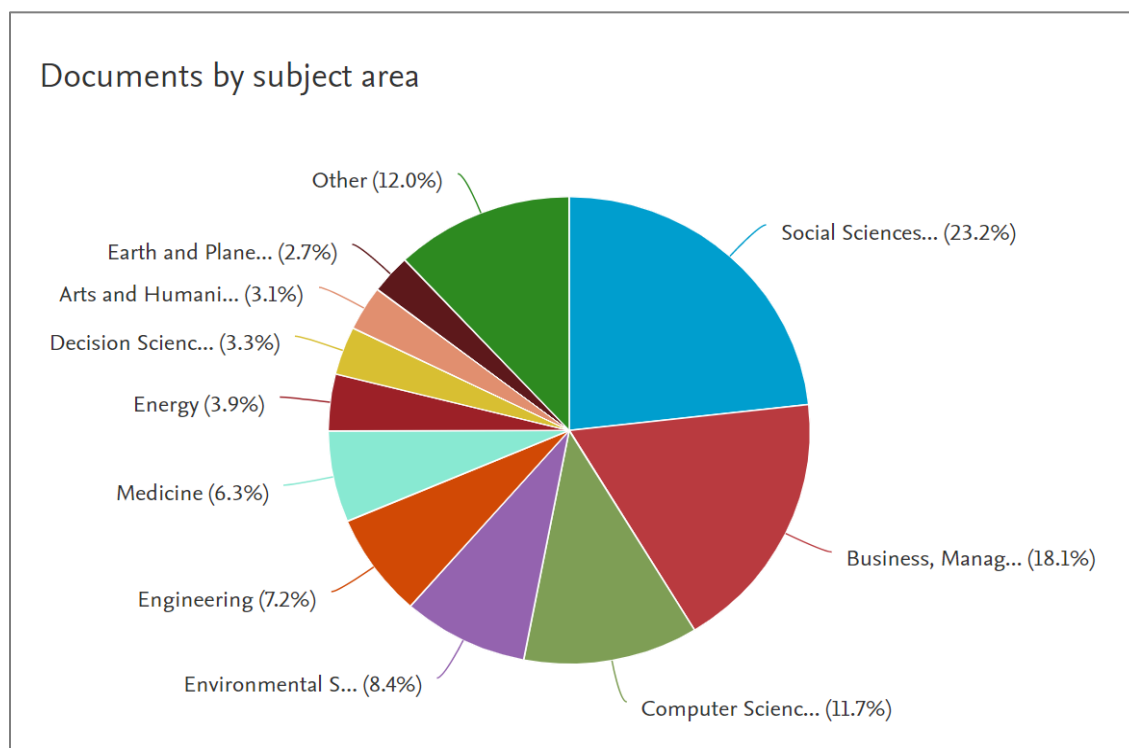


Figure 4. Distribution of sentiment analysis documents by different subject area

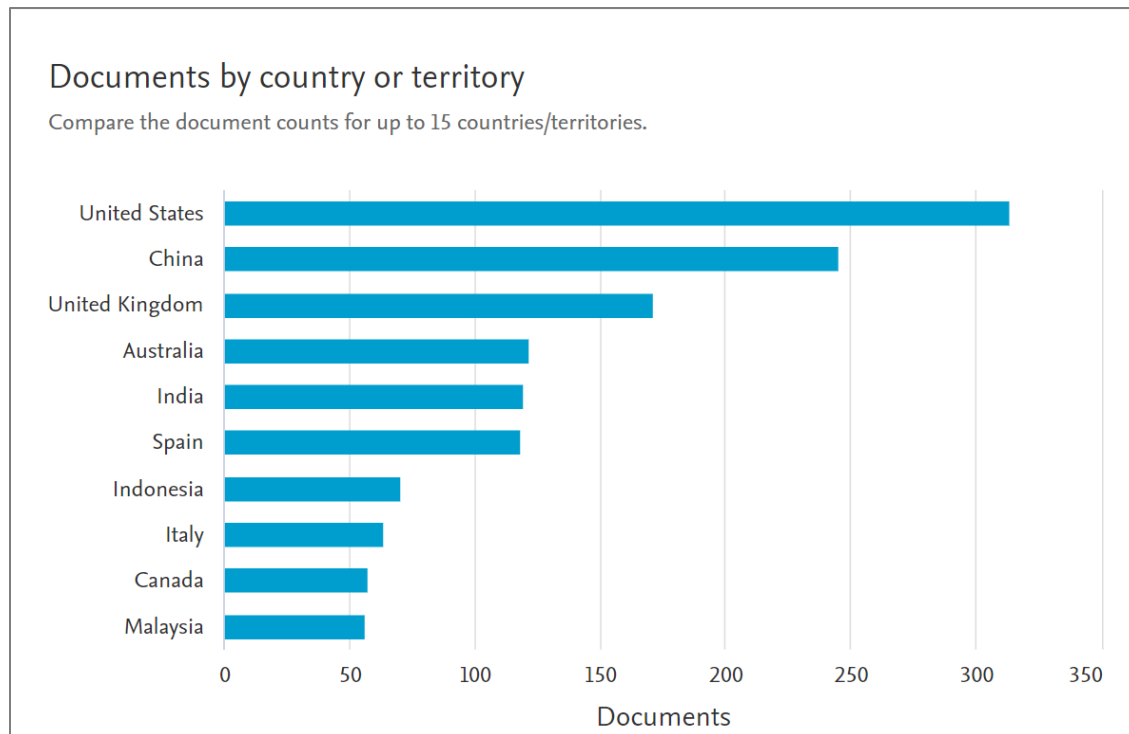


Figure 5. Distribution of published documents by country

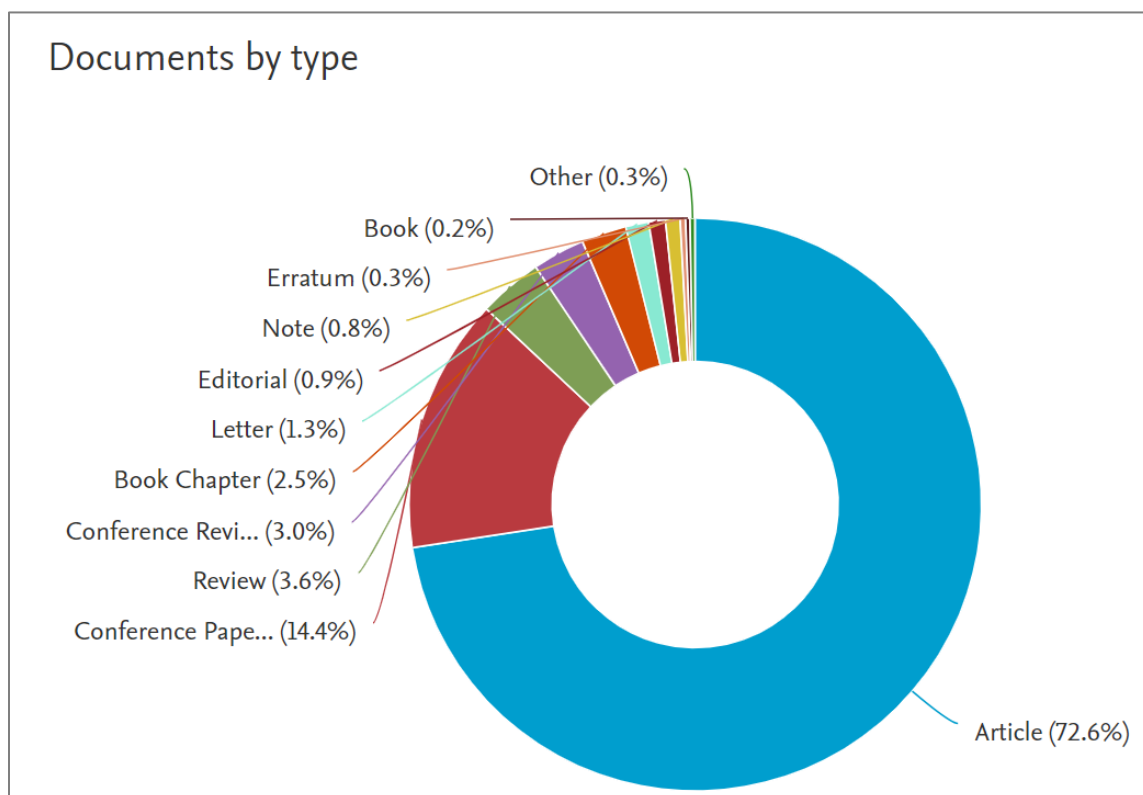


Figure 6. Distribution of published documents by type



Figure 7. Keyword cloud distribution

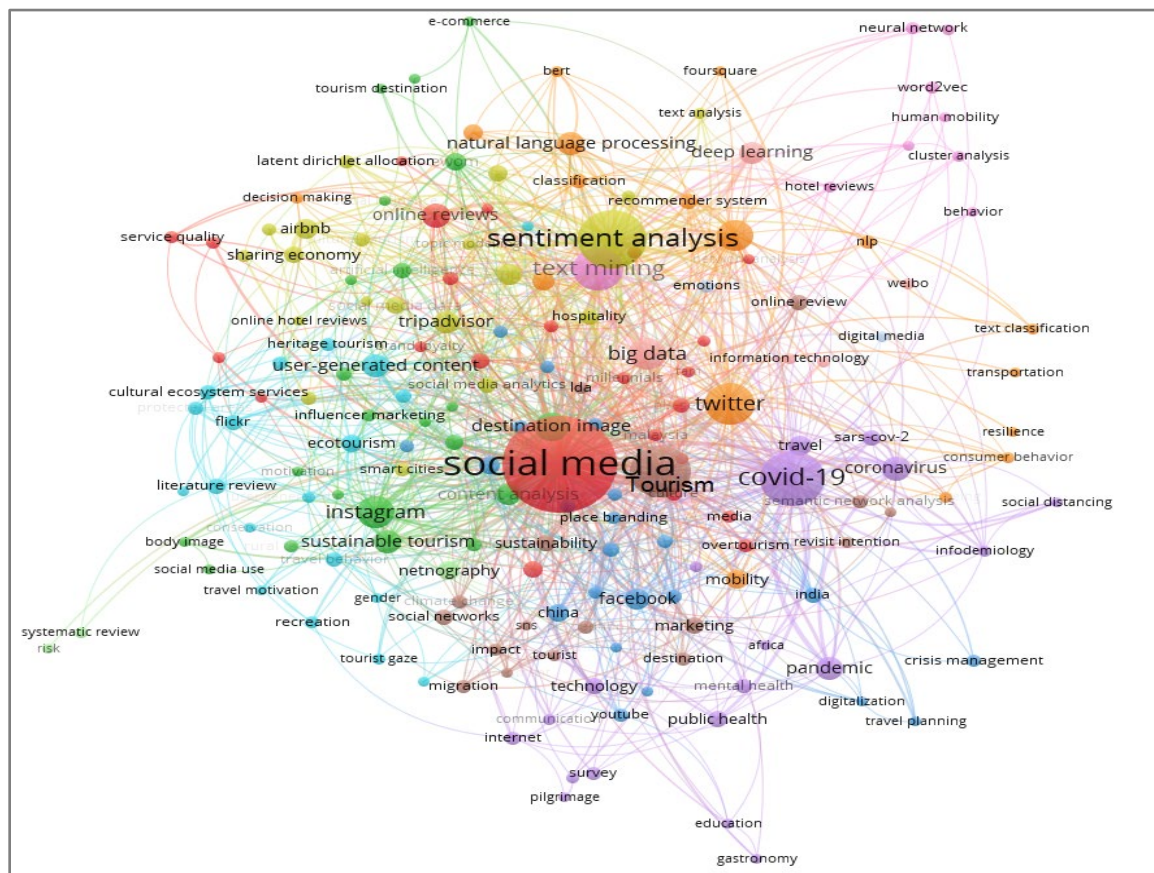


Figure 8. Network analysis with author keywords

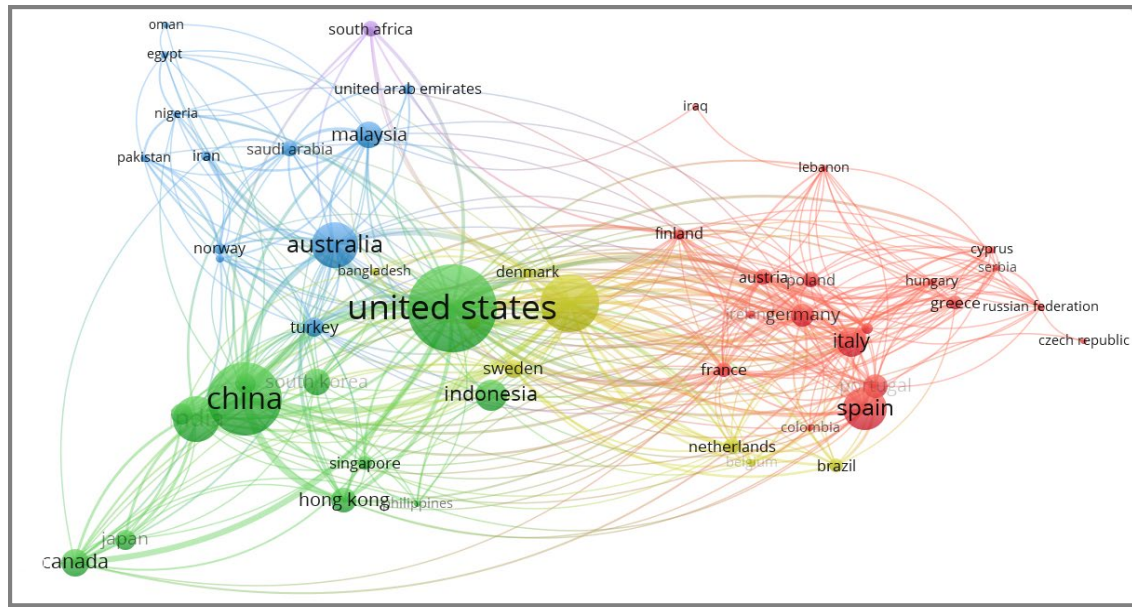


Figure 9. Network distribution of country co-authorship

5. Conclusion

This article provided a perspective on the global trends in 'sentiment analysis in the tourism industry during covid-19 pandemic' research. We performed a bibliometric analysis of scientific documents, subject categories' distribution, publications' venues, and geographic distribution. Our analysis shows that in the pandemic years, the research has been done in various subject areas, but 'Social Sciences,' 'Business Management' and 'Computer science' are the most frequent subject categories. There has been significant growth in the field. A small group of productive countries contributed to substantial articles: the top 10 countries contributed to ~77% of the total documents. The United States has the leading position in global sentiment analysis research. Similar to other research studies, this study is restricted in some directions. First, the data for the study is collected from the Scopus database. In future studies, other well-known databases could be included to collect data. Future studies can also analyze tourism-related books and other publications. We hope that this study will be useful for researchers trying to do targeted research in the sentiment analysis field for different domains and understanding the different aspects such as significant publications type, countries and keywords that have been used during the period of COVID-19.

Acknowledgments

The authors gratefully acknowledge the support of Effat College of Engineering at Effat University, Jeddah, KSA.

References

- Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., Almahdi, E. M., Chyad, M. A., Tareq, Z., Albahri, A. S., Hameed, H., and Alaa, M. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, 167, 114155-114155, 2021.
- B, Liu.. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. . Cambridge: Cambridge University Press. (2015)
- Chandra, R., and Krishna, A. COVID-19 sentiment analysis via deep learning during the rise of novel cases. *PloS One*, 16(8), e0255615, 2021.
- Cretu, C.-M., Turtureanu, A.-G., Sirbu, C.-G., Chitu, F., and Marinescu, E. , Tourists' Perceptions Regarding Traveling for Recreational or Leisure Purposes in Time of Crisis. *MDPI*, 2021.
- Demola, O., Oluwaseun, K., Funmi, O., Adebawale, O., and Oluwasoye, M . Covid-19 and the tourism industry: An early-stage sentiment analysis of the impact of social media and stakeholder communication. *International Journal of Information Management Data Insights*, 2021

- Devi, C. N., and D. R.. Literature Review on Sentiment Analysis in Tourism. *TEST Engineering and Management* 2020.
- Flores-Ruiz, D., Elizondo-Salto, A., and María de la O Barroso-González. Using social media in tourist sentiment analysis: A case study of andalusia during the covid-19 pandemic. *Sustainability (Basel, Switzerland)*, 13(7), 3836, 2021.
- Goossens, C.. Tourism information and pleasure motivation. *Annals of Tourism Research*, 301-321.(2000)
- Hamidah, I., Sriyon, S., and Hudha, M. N. A Bibliometric Analysis of Covid-19 Research using VOSviewer. *Indonesian Journal of Science and Technology*, 2020.
- Kasper, W., and Vela, M. (n.d.). Sentiment Analysis for Hotel Reviews. *Computational Linguistics*, 45–52.
- Mitas, Ondrej and Yarnal, Careen & Chick, Garry. Jokes Build Community: Mature Tourists' Positive Emotions. *Annals of Tourism Research*. 39. 1884–1905. 10.1016/j.annals.2012.05.003, 2012.
- Mohamed Hussein, D. M.-D.A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, 330-338, 2018.
- Obembe, D., Kolade, O., Obembe, F., Owoseni, A., and Mafimisebi, O.. Covid-19 and the tourism industry: An early stage sentiment analysis of the impact of social media and stakeholder communication. *International Journal of Information Management Data Insights*, 1(2), 100040, 2021.
- Ondrej, K., and Antonin, S. Visual Analysis of Search Results in Scopus Database. *22nd International Conference on Theory and Practice of Digital Libraries*, 2018.
- Polyzos, S., Samitas, A., and Spyridou, A. Tourism demand and the COVID-19 pandemic: an LSTM approach. *Tourism Recreation Research*, 2020.
- Roman, M., Niedziółka, A., and Kranodebski, A. K. Respondents' Involvement in Tourist Activities at the Time of the COVID-19 Pandemic. *MDPI*, 2020.
- Schuckert, M., Xianwei, L., and Law, R.. Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel & Tourism Marketing*, 2015.
- Sigala, M. Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. 2021.
- Škare, M., Soriano, D., and Porada-Rochoń, M., Impact of COVID-19 on the travel and tourism industry. 2021
- Stathis Polyzos, A. S.. Tourism demand and the COVID-19 pandemic: an LSTM approach. *Tourism Recreation Research*. 2020
- Stylidis, D., and Terzidou, M. Tourism and the Economic Crisis in Kavala, Greece. *Annals of Tourism Research*. 2013.
- Stock, F. D. Tourism as a complex interdisciplinary research object. *Annals of Tourism Research*. (2011).
- Swapna, G., Shim, K. J., Jiang, A. W., and Lee, A. J. Exploring the Impact of COVID-19 on Aviation: A Text Mining Approach. *IEEE*, 2021.
- Thomas, D. C., and Devi, W. N. Role of Hotels in the Promotion of Tourism in Manipur- A Case Study of Select Hotels in Imphal City. *International Journal of Advance Research in Computer Science and Management Studies*, 2017.
- Vishal A. Kharde, S. S. Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, vol 139, no 11, April 2016, 2016.
- Yonggang Lu & Qiujie Zheng (2021) Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic, *Current Issues in Tourism*, 24:7, 892-898, 2021.

Biography

Dr. Tayeb Brahimi, Associate Professor at the College of Engineering, Effat University, Jeddah, Saudi Arabia, received his Ph.D. (1992) and Master's degree (1987) Ecole Polytechnique, University of Montreal, Canada. He has worked as Research Scientist under Bombardier Chair/Canadair from 1992-1998. In 1998, he joined Jeppesen DataPlan in California, then Peregrine System as a TS Analyst, Quality Assurance Engineer, and Consultant for Electronic data interchange (EDI) in Dallas, Texas. Dr. Tayeb Brahimi has been a consultant at IONPARA Inc. for wind energy and aeronautics. He published more than 100 articles in scientific journals and international conferences on renewable energy, aircraft icing, sustainability, artificial intelligence, and the use of technology to support learning. Among other activities, he is a reviewer for many international journals, invited speaker by the Japan Society of Mechanical Engineering, the Gulf Educational Conference, and the Int. Conference on Eng. Education & Research. He also participated in Public Debate on Energy organized by the Government of Quebec, Canada. Current research interests are related to renewable energy (solar, wind, wave, and waste to energy), simulation, sustainability, artificial intelligence, machine learning, and engineering education.

Dr. Akila, Assistant Professor at the Computer Science Department and Dean for Graduate Studies and Research at Effat University. She holds a Bachelor of Engineering degree in Computer Science from Institut National d'Informatique in Algeria (1988), a Master of Science degree from the University of Quebec (Montreal) and INI (Algeria) in 1992, and a Ph.D. in Computer Science from Ecole Supérieure Nationale d'Informatique (ESI) in 2011. She is an IEEE Senior Member. She worked as a Research Assistant and Software Engineer in Canada (Montreal) and as a Senior Software Engineer and Senior Member of Technical Staff at Harbinger, a B2B company in Concord, California. She joined Effat University in 2002. She is certified as a Cisco Regional Academy Instructor since 2004 and worked as a Cisco Coordinator till 2011. She served as the Chair of the CS department and then from 2013 as a Dean for the College of Engineering till summer 2017. Dr. Akila was the main lead for the College of Engineering NCAAA accreditation (CS, ECE, IS programs). She also led the ABET international accreditation for the three programs (CS, ECE, IS programs). She served as a program reviewer for NCAAA external panel. Her main research interests are in artificial intelligence, knowledge management, communities of practice, machine learning, sentiment analysis, big data, and computer network security. Dr. Sarirete has vast experience in the software development industry and software engineering. She is also interested in engineering education smart cities innovation, especially the human aspect and the collaborative work.

Asra Yaqub is a computer science undergraduate student at Effat University in Jeddah, Saudi Arabia. Her research interests revolve around the use of artificial intelligence and machine learning in the tourism industry. She worked on an augmented reality project for a multinational company in the aerospace sector and her current research interest focuses on the use of artificial intelligence and machine learning in the aerospace and tourism industry.

Huda Thalib is a computer science undergraduate student at Effat University in Jeddah, Saudi Arabia. Her primary area of study is the application of artificial intelligence and machine learning techniques in healthcare, electronic health records, and tourism industry.