Preprint submitted to the Current Issues in Tourism (Taylor & Francis). You can find the published version at:
https://www.tandfonline.com/doi/abs/10.1080/13683500.2022.2095507

COVID-19 and Tourism: Extracting public attitudes

Mahdi Hajiabadi[a] and Hamed Vahdat-Nejad[a]* and Hamideh Hajiabadi[b]

[a]*Department of Computer Engineering, University of Birjand, Birjand, Iran*

[b]*Department of Computer Engineering, Birjand University of Technology, Birjand, Iran*

Hamed Vahdat-Nejad*: vahdatnejad@birjand.ac.ir

Abstract- Taking advantage of the users' posts on Twitter, we investigate the impact of COVID-19 on tourism in the early months of the epidemic. For this purpose, more than two million tweets published in the first months of the outbreak are analyzed. A comprehensive lexicon of keywords in the field of tourism, as well as international airlines, is collected and used for extracting tourism-related tweets. Employing a new model based on the RoBERTa language, we extract the sentiments of tweets for different countries. The results show differences in users' positive or negative views in different countries. While in some countries, such as Germany, the public view is positive, the public view is negative in other countries, such as Russia.

Keywords: Tourism, Sentiment Analysis, NLP, Knowledge extraction, COVID-19

1. Introduction

Tourism is one of the most important industries that has been severely affected by COVID-19 epidemic constraints. The closure of tourist attractions, the restriction of flights, and the ban on tourists travel have caused enormous economic losses in tourism (Gössling et al., 2020). In this regard, numerous users published their ideas and observations regarding tourism on social networks. Users' reviews generally affect other users (Sparks et al., 2013) and provide rich resources of implicit information (Leung et al., 2013). Processing this huge amount of data is beneficial for organizations that care about the users' sentiments regarding their services and aim to reduce the negative effects of COVID-19 on the tourism industry.

Motivated by previous research (Lu & Zheng, 2021) (Sontayasara et al., 2021), we investigate users' viewpoints regarding the impact of COVID-19 on the tourism industry in the initial stage of the outbreak. We analyze more than two million English tweets related to COVID-19 between March 23 and June 23, 2020. To identify tweets related to tourism, a lexicon-based method is proposed. To find out the sentiment of each tweet, we employ the RoBERTa (Liu et al., 2019) language processing model. Finally, the trend of positive and negative sentiments is investigated for several countries, which are among the top 10 tourist destinations globally. Results show that the trend of negative tourism-related tweets correlates highly with the official statistics of COVID-19 infections in many countries. Besides, while in some countries like Germany and Japan, users were mostly positive regarding COVID 19, users were mostly negative in some other countries such as China and Russia.

The paper is organized as follows: Section 2 describes the background. Section 3 elaborates on the proposed method. The experiments are presented in Section 4, and Section 5 concludes the paper.

2. Background

Users publish their opinions and observations on various topics in comments and posts on social networks such as Twitter. Massive processing of users' comments leads to the extraction of high-level knowledge such as tourists' food preferences (Asani et al., 2020), tourists' visiting preferences (Abbasi-Moud et al., 2019; Abbasi-Moud et al., 2021), and tourists' emotions (Kumar & Zymbler, 2019).

Regarding the COVID-19 epidemic, Sari et al. (Sari & Ruldeviyani, 2020) investigated 340 tweets of public transport passengers in Indonesia and divided them into three categories including positive, negative, and neutral. The results show that most users were hoping to control the spread of the COVID-19 virus. Lu et al. analyzed 53,000 tweets related to cruise tourism from February 1 to June 18, 2020 (Lu & Zheng, 2021). They presented the average daily sentiment score trend, which shows that people are increasingly interested in traveling and exploring. Sontayasara et al. examined the views of people planning to travel to Bangkok by analyzing 881 English-language tweets and provided the number of positive, negative, and neutral tweets (Sontayasara et al., 2021).

To the best of our knowledge, none of the previous studies have examined the impact of COVID-19 on tourism in the world. Continuing the past research, this article analyzes the sentiment of the tweets related to tourism for the global top tourism destinations.

3. Proposed method

Our initial dataset includes two million COVID-19-related tweets for 14 weeks from March 23 to June 23, 2020. We propose a lexicon containing keywords related to tourism to extract tourism-related tweets. For this purpose, we extract the words related to tourism from the Oxford[1] dictionary and the Wordnet[2] ontology. Then, irrelevant words are removed. We also add the keywords of all international airlines to the dictionary. Finally, our proposed dictionary contains 77 keywords related to tourism, such as tourism, accommodation, and travel, and 956 keywords related to airlines[3] such as Lufthansa and Victory Air Transport.

---

[1] https://www.oxfordreference.com

[2] https://wordnet.princeton.edu

[3] https://en.wikipedia.org/wiki/Lists_of_airlines#By_continent

Afterward, each tweet's words are compared with the whole lexicon keywords. If one of the keywords is repeated in the tweet's text, that tweet is extracted for the tourism dataset. In summary, the number of tourism tweets extracted is 27112 tweets (average 2000 tweets per week). Then, the location of each tweet is identified using our previous content analysis method (Salmani et al., 2021).

Sentiment analysis is a well-known process that detects the sentiment of a tweet and helps understand the changing trend of public opinion about tourism during the outbreak. We use RoBERTa (Liu et al., 2019) language model (i.e., an optimization of the BERT method (Devlin et al., 2019)), which has high accuracy (Ghasiya & Okamura, 2021), to determine the sentiment (neutral, positive, or negative) of the tweets. RoBERTa provides a set of token representations: h_t ($t = 1,...,T$) for each tweet. For sentiment detection, it uses three classifiers trained with three datasets, including the Stanford Sentiment treebank (Socher et al., 2013) (This dataset includes experts reviews about movies and contains 5154 phrases, of which 67300 samples have been used), SemEval 2015 Task 10 (Liu et al., 2019) (This dataset contains general tweets from January 2012 to January 2013, in which 6800 samples were used), and SemEval 2015 Task 11 (Ghosh et al., 2015) (This dataset is about irony, sarcasm, and metaphor, in which 3500 samples are used). Finally, the sentiment of each tweet is identified using a majority vote between the three classifiers.

After performing the above processes, a neutral, positive, or negative label is assigned to each tweet. The following are examples of positive and negative tweets:

- Positive: "We've just about beaten Covid as well. We will hopefully be able to travel soon."
- Negative: "Corona Virus ruined my trip to Jamaica I was leaving Monday 😭."

Afterward, the number of positive and negative tweets for each week is obtained, which reflects the public sentiment changes over time. Moreover, the percentage of positive and negative tweets is calculated for each country. Users' sentiments reveal implicit information about the impact of COVID-19 on tourism from the users' point of view.

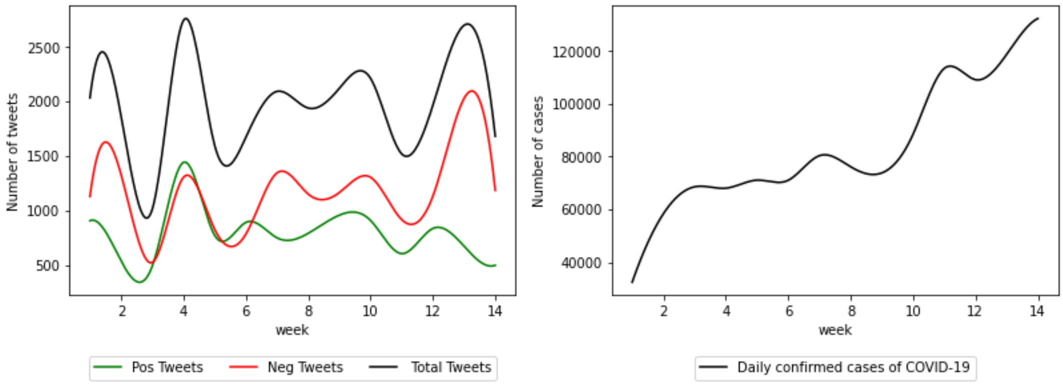
4. Experiment

We have used the GATE[4] (Cunningham, 2002) open-source tool to implement the steps of the proposed method. GATE is a multilingual platform generally used for text engineering (Cunningham et al., 2014). In this platform, a set of tools and capabilities required for text processing are available (Digan et al., 2021). This tool increases the accuracy of tweets annotation (for location and tourism) by pre-processing the text of tweets.

For analysis, we select 19 countries with the highest number of confirmed cases of COVID-19 as well as the number of tourism-related tweets. Figure 1 (a) shows the total as well as the number of positive and negative tourism-related tweets weekly from March 23 to June 23, 2020. Figure 1 (b) shows the trend of the number of daily confirmed cases of COVID-19 in the world in the investigated weeks based on the official statistics reported by the World Health Organization. The total number of tweets curve shows a relative equilibrium. Over time, the number of negative tweets becomes relatively more than the number of positive tweets. It indicates that users' perceptions of tourism were becoming more gloomy day by day as the number of COVID-19 confirmed cases was increasing. From week six onwards, the average number of negative tweets is much

[4] https://gate.ac.uk/

higher than the number of positive tweets, which might be attributed to travel restrictions due to the spread of the virus in these weeks.



(a) (b)

Figure 1. (a) Frequency of tourism-related tweets and (b) The official confirmed cases of COVID-19 in the world

89% of the total tourist tweets are related to 9 countries, including (in descending order) India, USA, China, UK, Australia, Pakistan, Japan, France, and Germany, and 11% of the total tweets are related to 10 countries, including (in descending order) Brazil, Italy, Russia, UAE, Canada, Saudi Arabia, Qatar, Singapore, Chile, and Iran. Figure 2 shows the cumulative percentage of sentiments for each of these countries. The appearance order of countries is based on the number of tourism tweets.

These countries can be divided into three main categories:

- Countries with a positive sentiment in which users' attitude is averagely positive, including Germany, Japan, Singapore, and the UAE. Users in these countries have mainly expressed satisfaction with the resumption of domestic flight services and satisfaction with the hygienic principles in travel.

- Countries almost dominated by negative sentiments, including China, the UK, Russia, Brazil, Canada, Qatar, Saudi Arabia, and Chile. Users in these countries have mostly tweeted about the closure of leisure centers, travel bans, and the lack of hygiene by travelers and travel officials such as flight attendants and hotel staff.

Countries with an equal number of positive and negative sentiments in which users' attitudes are somehow neutral on average, including India, the USA, Australia, Pakistan, France, Italy, and Iran.

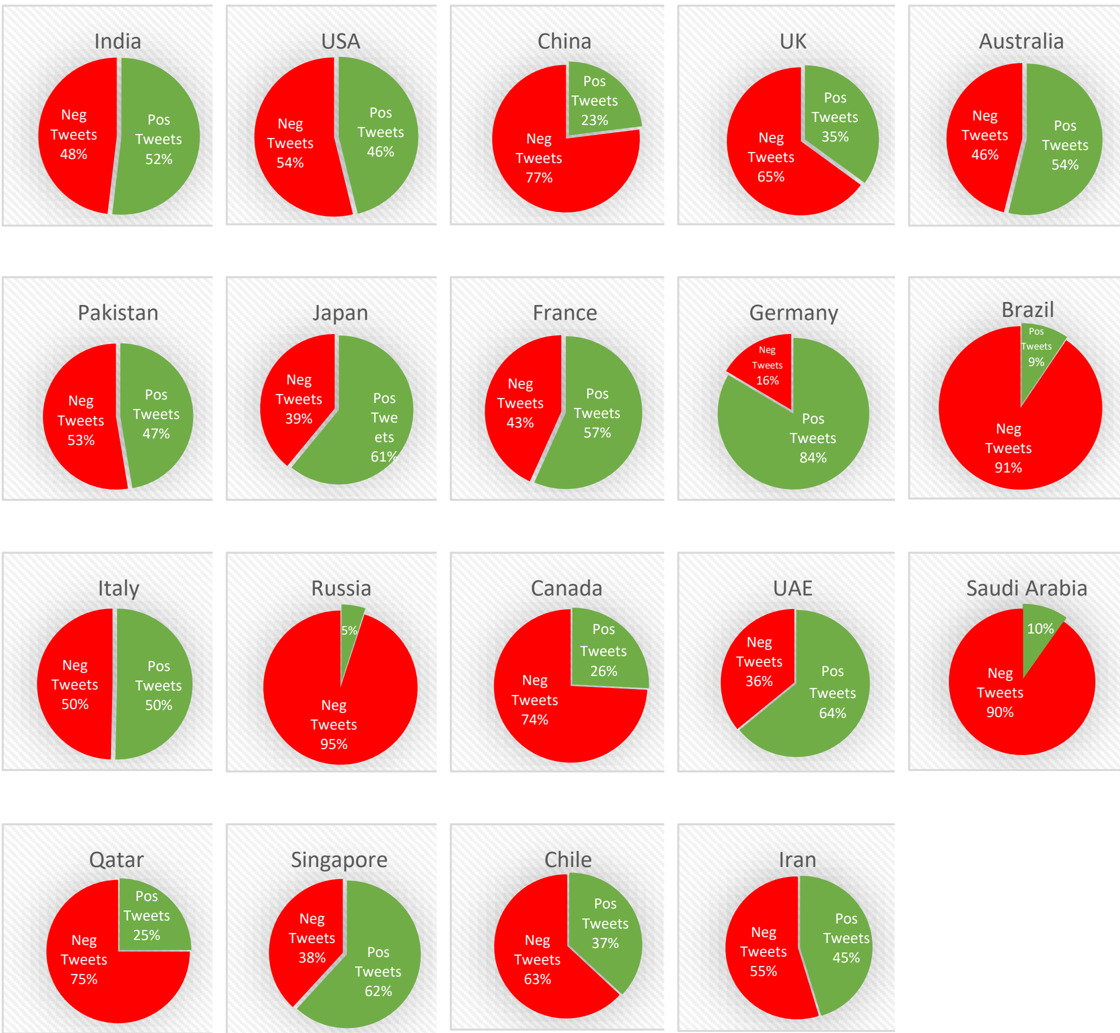


Figure 2. Cumulative sentiment analysis

Tweets sentiments are different in these countries for several reasons: (a) different statistics of patients with the disease, (b) different COVID-19 control policies (such as enforcing masking and lockdown, flights cancelation, etc.), (c) the operating situation of tourism-related businesses (such as transportation and accommodation), and (d) certain news and events.

Figure 3 shows the weekly positive and negative tweets percentage for the nine countries with the most tweet proportion. Most of these countries are the major destinations for tourists. Other countries do not have enough weekly tweets to be analyzed.

Analyzing the diagrams in some countries such as the USA, Japan, Australia, and France reveals a correlation between the trend of the COVID-19 statistics and the percentage of negative tweets. Special news and events have also influenced the charts. For example, in the first weeks of the epidemic, although the number of cases in India was insignificant, the number of negative tweets is very high because of the non-closure of Delhi airport and the non-observance of hygienic principles. Also, in China, in the 9th to 11th weeks, despite the low number of official cases, the percentage of negative tweets has increased significantly due to President Trump's announcement of the travel ban from the United States to China. Also, in the 13th week, the percentage of negative tweets reaches its peak due to users' anger over the cancellation of more than a thousand flights at Beijing airport.

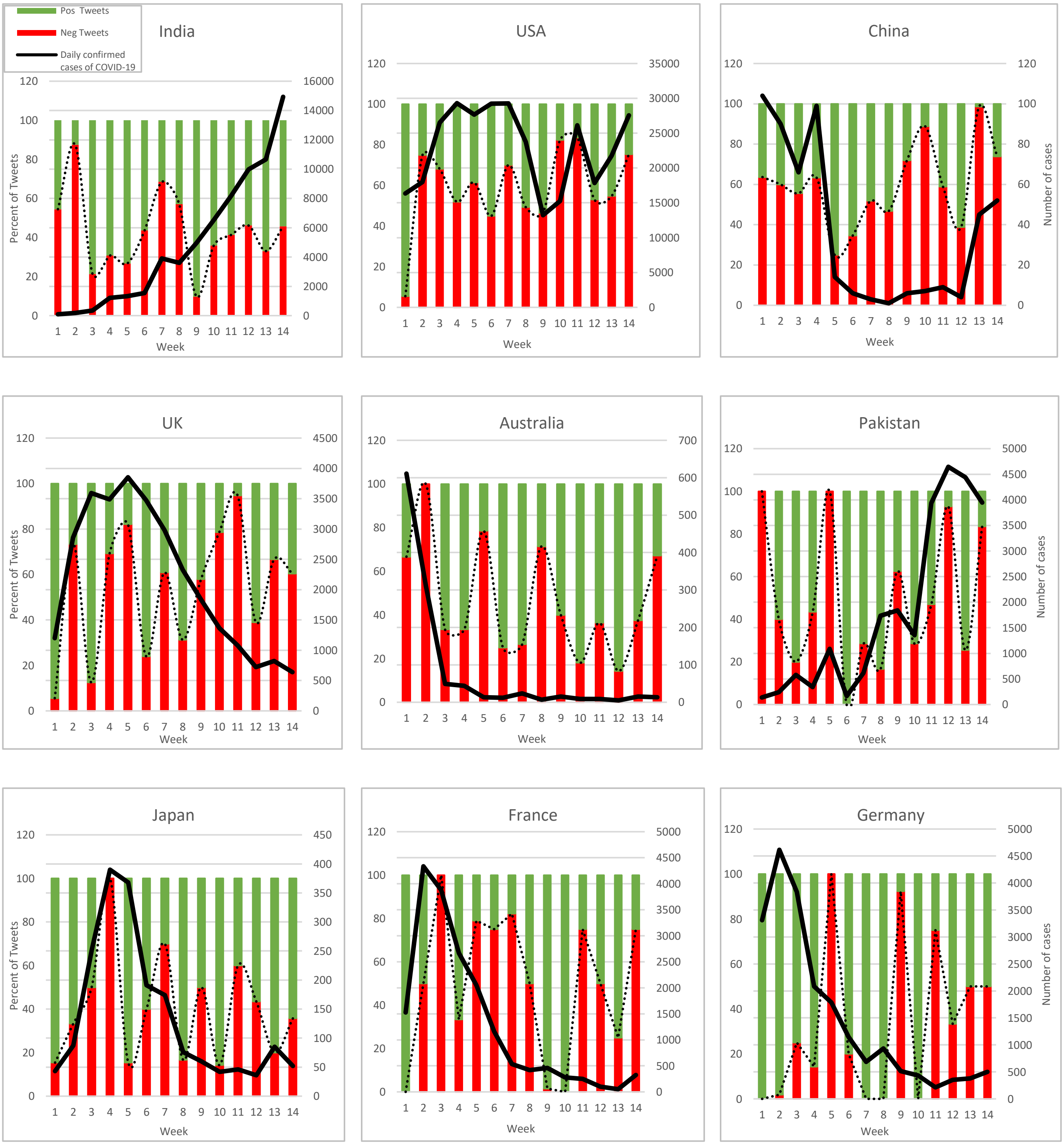


Figure 3. Percentage of negative vs. positive sentiments

In the United Kingdom diagram, the number of negative tweets rises during the 8th to 11th weeks, while the number of patients decreases. This is due to the closure of hotels, the cancellation of trips, the impact of the tourism decline on the royal budget, and quarantine requests. Also, in Germany, the number of negative tweets in the 11th week peaked due to the billion-dollar loss of the German airline Lufthansa.

5. Conclusion

In this research, the impact of COVID-19 on tourism has been explored by processing COVID-19-related tweets. To extract tourism-related tweets, a comprehensive lexicon including the words related to tourism and international airlines has been proposed. Then, using a pipeline in GATE software, related tweets were extracted and geographically analyzed. COVID-19 was a global epidemic; however, different countries have had different attitudes. This is due to differences in the culture, outbreak patterns of the disease, special news and events, the type of health instructions, and governments' policies. For example, positive sentiments have prevailed in some countries, such as Germany, while negative sentiments have prevailed in some countries, such as Russia, Brazil, and Saudi Arabia.

To continue, tourism can be studied by focusing on important aspects such as economy, attractions, tourists' satisfaction, etc. Besides, identifying the main topics of users reviews is another future research direction.

Disclosure statement

The authors report there is no competing interest to declare.

Reference:

Abbasi-Moud, Z., Vahdat-Nejad, H., & Mansoor, W. (2019). Detecting Tourist's Preferences by Sentiment Analysis in Smart Cities. Global Conference on Internet of Things, Dubai, United Arab Emirates.

Abbasi-Moud, Z., Vahdat-Nejad, H., & Sadri, J. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications*, *167*, 114324. https://doi.org/https://doi.org/10.1016/j.eswa.2020.114324

Asani, E., Vahdat-nejad, H., Hosseinabadi, S., & Sadri, J. (2020). Extracting User's Food Preferences by Sentiment Analysis. 8th Iranian joint congress on fuzzy and intelligent systems Mashhad, Iran.

Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities*, *36*(2), 223-254. https://doi.org/https://doi.org/10.1023/A:1014348124664

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., & et al. (2014). *Developing Language Processing Components with GATE* (8 ed.). University of Sheffield, UK. http://gate.ac.uk/sale/tao/index.html

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, USA.

Digan, W., Névéol, A., Neuraz, A., Wack, M., Baudoin, D., & et al. (2021). Can reproducibility be improved in clinical natural language processing? A study of 7 clinical NLP suites. *Journal of the American Medical Informatics Association*, *28*(3), 504-515. https://doi.org/https://doi.org/10.1093/jamia/ocaa261

Ghasiya, P., & Okamura, K. (2021). Investigating covid-19 news across four nations: A topic modeling and sentiment analysis approach. *Ieee Access*, *9*, 36645-36656. https://doi.org/https://doi.org/10.1109/ACCESS.2021.3062875

Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., & et al. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. 9th international workshop on semantic evaluation, Denver, Colorado.

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism*, *29*(1), 1-20. https://doi.org/https://doi.org/10.1080/09669582.2020.1758708

Kumar, S., & Zymbler, M. (2019). A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data*, *6*, 1-16. https://doi.org/http://dx.doi.org/10.1186/s40537-019-0224-1

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of travel & tourism marketing*, *30*(1-2), 3-22. https://doi.org/https://doi.org/10.1080/10548408.2013.750919

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., & et al. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint*, 1-13. https://doi.org/https://doi.org/10.48550/arXiv.1907.11692

Lu, Y., & Zheng, Q. (2021). Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic. *Current Issues in Tourism*, *24*(7), 892-898. https://doi.org/https://doi.org/10.1080/13683500.2020.1843607

Salmani, F., Vahdat-Nejad, H., & Hajiabadi, H. (2021). Analyzing the Impact of COVID-19 on Economy from the Perspective of Users Reviews. 11th International Conference on Computer Engineering and Knowledge, Mashhad, Iran.

Sari, I. C., & Ruldeviyani, Y. (2020). Sentiment analysis of the covid-19 virus infection in indonesian public transportation on twitter data: A case study of commuter line passengers. 2020 International Workshop on Big Data and Information Security (IWBIS), Depok, Indonesia.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., & et al. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. conference on empirical methods in natural language processing, Seattle, Washington, USA.

Sontayasara, T., Jariyapongpaiboon, S., Promjun, A., Seelpipat, N., Saengtabtim, K., & et al. (2021). Twitter Sentiment Analysis of Bangkok Tourism During COVID-19 Pandemic Using Support Vector Machine Algorithm. *Journal of Disaster Research*, *16*(1), 24-30. https://doi.org/https://doi.org/10.20965/jdr.2021.p0024

Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, *39*, 1-9. https://doi.org/https://doi.org/10.1016/j.tourman.2013.03.007