Twitter Opinion Analysis about 5G Technology

Mukhtiar Singh
Ph.D. Research Scholar,
Department of Computer Science, Punjabi University, Patiala.
Email: mukhtiarrai73@gmail.com

Abstract: Since thousands of users freely express their opinions on Twitter every day, it has become a rich source for sentiment analysis and opinion mining data. In this investigation, we look at the sentiment of shared articles with the hashtag "#5G" and categories it as positive, negative, or neutral. We used statistical sentiment analysis tool to create a classification model that had an accuracy and recall of 83.69%. The findings indicate that it is possible to recognize key public opinion factors in the acceptance or rejection of 5G technology, which is valuable information for technology companies.
Keywords: Twitter, Sentiment analysis, 5G technology.

1. INTRODUCTION

So after the 4G technology, 5G is the next generation of cellular phone technologies. It is not standardized at this time, and telecommunication companies are working on prototypes. The first functional devices are expected to be available in 2020, mainly in 4G-enabled cities (Gohil et al., 2013) . It will be able to navigate through a mobile device at 400 megabit speeds due to the 5G network. The advantages, according to (Go et al., 2009), include high capacity and speed, global connectivity and service portability, high resolution and bi-directional broad bandwidth, and 5G technology acting as a backbone for multimedia, speech, and the internet.

Disadvantages: getting a high speed in certain parts of the world is challenging, security and privacy issues in 5G need to be addressed, and old devices need to be upgraded to support 5G technology (Agarwal et al., 2019). In recent years, a large number of people have switched to social media sites like Twitter to express their thoughts on various things, locations, and topics.

Recent studies have also shown that, in comparison to conventional methods, Twitter can provide plausible details about opinions (Filho et al., 2015) and (Hoffman et al., 2010).

The aim of this study is to analyze the opinions of Twitter users who used the hashtag #5G in order to determine how people are behaving before the technology became popular. The hashtags were used to get the data from Twitter's official API. Following the preprocessing of the data, we conduct a sentiment analysis using statistical sentiment analysis tool to fulfill the final tasks. Such a tool was used for word clustering and identifying the main themes about the 5G network, allowing you to better understand the data. The below is just how the rest of the article is coordinated: The materials used and the processes used are listed in Section 3, the findings are shown in Section 4, and the debates and conclusions are discussed in Sections 5 and 6.

2. RELATED WORK

This section represents problems relating to sentiment analysis on Twitter, such as analysing and determining the polarity of tweets using machine learning techniques and text mining tools. (Loyola-Gonzalez et al., 2019) proposed a contrast pattern-based classifier for detecting Twitter bots. Their purpose was to support the expert in taking legal action against an account that had displayed suspicious behavior. They investigate the use of Bayesian Network, k-Nearest Neighbor, Naive Bayes, Random Forest, Support Vector Machine, Logistic Regression, Adaptive Boosting, Bagging, and Multilayer Perceptron in their work. They wanted to find the best classifier for bot detection. According to their findings, the best classifier was Random Forest, which had a 99% Area Under the Curve (AUC) and a Matthews Correlation Coefficient (MCC). While the pattern-based classifier only achieved 90% AUC and 91% MCC (Loyola-Gonzalez et al., 2019).

(Martin-Domingo et al., 2019) used sentiment analysis to analyse opinions on the Londres Heathrow Twitter account in order to identify new ideas that could improve airport service quality (ASQ). They developed a system for collecting and analysing data. Their design included a list of 128 key words for identifying tweets that discuss the quality of tweets about the ASQ. Their sample size was 5,332 tweets. Their research concluded that the ASQ possessed 23 attributes. They performed the sentiment analysis using the tools. They Say and Twinword, and

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

they manually verified the results of both tools. They were able to achieve 78.7% and 69.6% accuracy respectively, giving them the best performance (Martin-Domingo et al., 2019).

Using machine learning, (Reyes-Menendez et al., 2018), demonstrated a sentiment analysis based on the #BlackFriday Twitter hashtag. This was the goal of identifying the consumer's identity and sentiment in relation to the three classes: positive, negative, and neutral. They were attempting to connect this sentiment to the offers made available by various companies. They collected a total of 2,204 tweets using the Twitter API, and the companies that were analysed were chosen because they are on the “El Economista” national list of companies. They used the online tool Monkey learns, and as a result, 32% of the tweets were classified as positive, 7.7% as negative, and 60.25% as neutral. Their results are aimed to create better marketing strategies (Saura et al., 2018).

(Mane et al., 2014) proposed a method for performing sentiment analysis on large amounts of unstructured data using a Hadoop Cluster. They were used a distributed environment. Their main contribution is based on speed, but they also used three sentiment analysis classes: positive, negative, and neutral. They used data from the streaming API to collect tweets in real-time, SetiWordNet to create a word dictionary, and the information codified in WordNet to assign a sentiment analysis score. They, on the other hand, preprocessed and labelled the data using the OpenNLP tool. On the three sentiment classes, they reported an accuracy of 72.27% (Mane et al., 2014).

(H. Wang et al., 2012) the opinions of three large pizza chains: Pizza Hut, Domino's Pizza, and Papa John's Pizza, with the goal of using the knowledge of the opinions to improve the companies' products and services. The authors collected data from social networks and analysed the data using mining techniques. They begin by preprocessing the data in order to structure the information and integrate the obtained data, before transforming the obtained data into a usable format. They extracted, clustered, and indexed the data using the SPSS Clementine tool. They also used Nvivo 9 to run queries and look for patterns in the data. The results show that Domino's Pizza is more involved with their customers, as evidenced by the number of posts and responses. This demonstrated that businesses should maintain an open dialogue with their

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

customers, monitor the conversation, and respond to their customers' complaints and concerns in a timely manner in order to avoid crises caused by price increases, recipe changes, and so on.

(Chamlertwat et al., 2012) used machine learning to perform sentiment analysis on Twitter about some smartphone characteristics. This is to discover the consumers' perceptions and classify the comments into two categories: positive or negative. They devised a system that was divided into stages to accomplish this. The first stage was in charge of gathering data through the Twitter API. They proposed a set of keywords for collecting tweets about the most popular products on the market. In the second stage, they filter those tweets that were considered opinion, for which they used manual labelling and developed an SVM model to learn to distinguish between opinion and not. They used a variety of information gain evaluations and achieved an accuracy of 84.5 %. In stage three, they used the SentiWordNet tool to determine the polarity of the tweets. Finally, in the fourth stage, they classified product characteristics. To validate their methodology, they compared the identified product characteristics with those obtained by experts; there were similarities between both analyses. The authors emphasise that these types of studies provide information that assists industry in making decisions and improving their products; they also mention that the more data collected, the better the results that can be achieved (Chamlertwat et al., 2012).

(W. Wang & Wu, 2011) investigated a system for real-time Twitter sentiment analysis for presidential candidates in the 2012 United States elections, with the goal of quantifying public sentiment toward the candidates and electoral events.

The algorithm used was a Naive Bayes with unigrams. Based on the current findings, it was possible to achieve 59% classification precision in categories such as negative, positive, neutral, and unknown (H. Wang et al., 2012).

(Samuels & McGonical, 2020) proposed a combination of techniques and algorithms for sentiment analysis in Twitter, including unigrams, senti-features, and tree kernels. They used an SVM model in a cross validation setting to validate the performance of each algorithm. The results show that for a binary classification of polarity, positive vs. negative, the Three Kernel model performs better with an accuracy of 73.93%, whereas the combination of Unigram + Senti-features incrementally improves the performance up to 75.39% (Agarwal et al., 2019).

(Go et al., 2009) proposed a machine learning technique that uses distant supervising to create positive and negative binary class for tweet classification. They comparison the Naïve Bayes, MaxEnt, and SVM models with the Unigram, Bigram, Unigram + Bigram, and Unigram + POS features. They discovered that using Unigram was the best condition when using an SVM model, with a performance of 82.2% accuracy, while Bigram with NBayes achieved 81.6% accuracy. Furthermore, the combination of Unigram + Bigram using MaxEnt has an accuracy of 83.0%, while Unigram + Tags POS has an accuracy of 81.9%. They noted that using emojis helped them achieve a score of more than 80% (Go et al., 2009).

3. MATERIAL AND METHOD

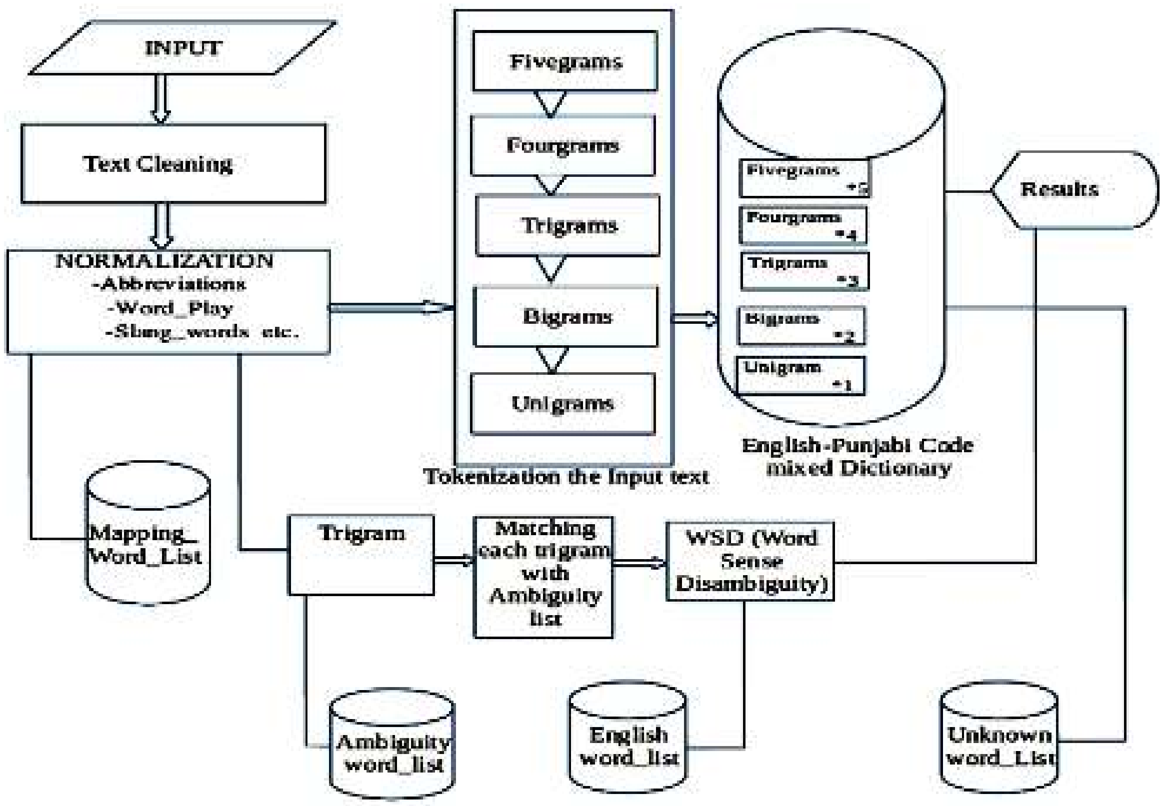

Figure 1: Development of the methodological process

Below figure 1 depicts the methodological process used in this study. The first involves collecting data from Twitter users by searching for the #5G hashtag with Python as a

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

programming language and connecting to the Twitter Search API. Step two involves categorizing tweets into three categories: positive, negative, and neutral. After that it is divided into four major steps as follow:

a) Input Phase

In this phase, the polarity of sentence entered by the user is tested. Initially, a sentence written in English-Punjabi code-mixed language is given to system. After input phase, the input text is cleaned in the Text_Cleaning process and then normalization and tokenization tasks are done which are explained below:

b) Normalization

As it has been analyzed by us, punctuation marks do not affect the polarity of a sentence. So, before processing the text we need to normalize the text data. In this phase, all the punctuation marks are found and deleted. The abbreviations or slang words are also removed by Mapping_Word_List. After normalization, the given input text is tokenized using N-gram approach.

c) Tokenization

First of all, the input sentence is tokenized in N-gram. After each sentence is tokenized, it is further split into N-gram tokens using the delimiter space (" ").

d) Finding Polarity

Initially, for all the N-gram tokens in dictionary, 0 frequency is assigned for the positive and negative polarity. Using N-gram approach, the frequency of sentence is found up to Five-gram approach in case of both positive and negative polarity using following function:

$$\mathrm{POS}_{\mathrm{SCORE}_\mathrm{i}} = \frac{\mathrm{PF}_\mathrm{i}}{\mathrm{PF}_\mathrm{i} + \mathrm{NF}_i + 1}$$

$$\mathrm{NEG}_{\mathrm{SCORE}_\mathrm{i}} = \frac{\mathrm{NF}_\mathrm{i}}{\mathrm{PF}_\mathrm{i} + \mathrm{NF}_\mathrm{i} + 1}$$

Where,

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

"POS_{SCOREi}" and "NEG_{SCOREi}" are sentiment polarities, which are found in a sentence.
"PFi" is the positive sentiment frequency of N-gram sentiment tokens in senetence which is obtained from Positive_Dictionary.
"NFi" is the negative sentiment frequency of N-gram sentiment tokens in senetence which is obtained from Negative_Dictionary.
"POS_{SCOREi}" is the positive score of N-gram sentiment tokens in a sentence currently processed which is considered as total score of N-gram sentiment positive tokens over total score of the both N-gram sentiment positive and N-gram sentiment negative tokens in a sentence.
"NEG_{SOCREi}" is the negative score of N-gram sentiment words in a sentence currently processed which is considered as total score of the negative N-gram sentiment tokens over total score of the both positive and negative N-gram sentiment tokens in a sentence.
Where, increment by one (+1) is used to remove the "∞" (infinity), because if in any case, total number of positive and negative score comes out to be 0, then, their score will be infinity.

3.1 Twitter

Social media has changed the way people interact in recent years (Yu et al., 2013), primarily because it has become an ideal medium for sharing opinions, experiences, and accomplishments, among other things. With 326 million active users, Twitter is one of the most popular social networks in the world (*Global Digital Report: Digital in 2019*, n.d.). This has sparked a strong interest among researchers, businesses, and governments to analyse users' opinions on specific topics of interest, attempting to determine contextual polarity by categorising tweets into three categories: positive, negative, and neutral. One of Twitter's greatest advantages is that it allows users to create and share opinions instantly in a public, easy, and accessible manner. It also enables interaction among users who are sharing or interacting with the same hashtag about a specific topic. A hashtag is a label represented by the symbol "#." This label enables users to participate in a larger discussion about a specific topic (Small, 2011). In contrast to other platforms, Twitter's public openness makes it ideal for information recollection (Neethu & Rajasree, 2013).

3.2 Twitter API

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

We used the Tweepy library to collect tweets because it allows us to connect to the Twitter API. A limitation is that the library only returns messages from the last 8 days, but another feature is that it allows you to obtain messages longer than 140 characters, up to 280, which the new length is allowed by the platform. We set the parameter "retryonratelimit" to true, which causes consults to try again after reaching the platform's limits and allows us to collect data continuously. We also used the “max id” argument to restart a collection that had been stopped in the previous iteration. The obtained data is semi-structured, which allows for the implementation of filters and queries on the data. For example, you could filter the data by date, geographic area, or sort the users by number of posts or followers, among other things.

3.3 Sample

(Cavazos-Rehg et al., 2018) and (Bermingham & Smeaton, 2011)] used sentiment analysis on two samples of tweets, one with 5,000 tweets and the other with 7,203 tweets, in previous studies.(Reyes-Menendez et al., 2018), on the other hand, collected 5,873 tweets related to the #WorldEnvironmentDay hashtag. Based on these findings, we decided to collect 8,102 tweets containing the hashtag #5G. The sample originally consisted of 46,559 tweets, but after removing duplicates, mentions, and retweets; our collection now consists of 8,102 tweets. The sample was validated using the following factors, as described by (Saura et al., 2018) and Reyes-(Reyes-Menendez et al., 2018):

1. A profile picture and a cover photo are required for the user.
2. Because retweets are considered duplicate information, they were discarded.
3. The majority of tweets have a localization name. Localization of precise names was replaced by their country, for example, California USA was replaced by ISA.
4. Several aspects, such as hashtags and links, were cleaned as part of the cleaning preprocessing. Asterisks, parentheses, brackets, underscore, dash, tilde, apostrophe, single and double quotes, greater than and less than symbols, numbers, and whitespaces are also permitted.
5. Tweets must be at least 80 characters long.

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

6. Only tweets in English were taken into account. It is important to note that tweets come from all over the world. However, because the processing tools only support English, we only used those in English.

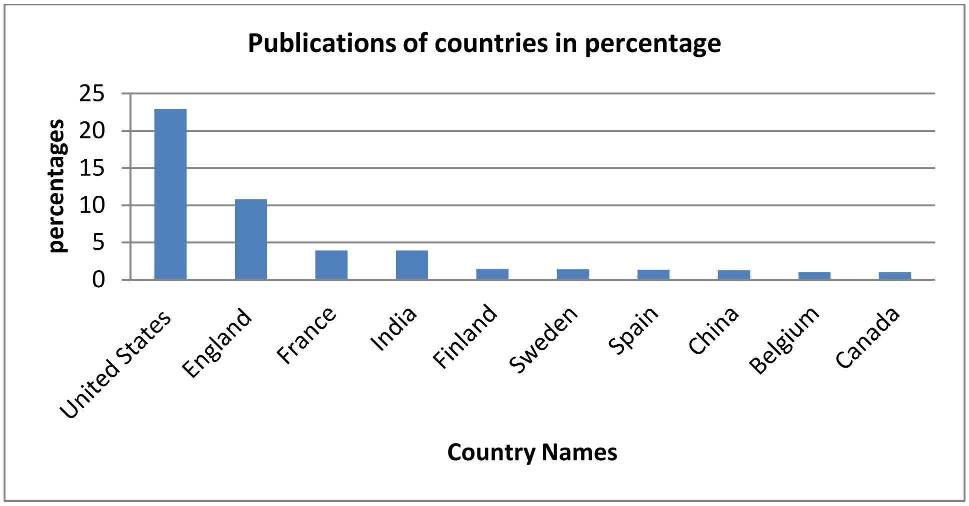

Figure 2: Publications of countries in percentage.

States, England, and France were the three most active countries. China, Belgium, and Canada were the three least active countries. It's understandable that they're coming from China, given that the platform isn't available.

3.4 Sentiment Analysis

The sentiment analysis method provides information about the user's mental state regarding a specific topic. This is that can be linked to social networks because it is not limited to them. According to (Liu, 2012), it is a field that analyses people's opinions, sentiments, evaluations, attitudes, and emotions using written text. Using machine learning techniques to analyse the emotional behavior's contained in the text is one of the most common ways to analyse the emotional behavior's contained in the text; machine learning is a subfield of computer science and artificial intelligence. Its inner workings are focused on data patterns for the creation of intelligent algorithms that can learn without relying on programming based on rules, but rather on behavior codified in examples (Holzinger, 2017).

Furthermore, the Google company has created a number of applications that have aided the work of many people in various aspects of their digital lives. Recently, the emphasis has been on the development of intelligent applications using artificial intelligence. Using statistical sentiment analysis, this tool can classify, extract, and detect sentiment.

4. RESULTS

This section presents the results obtained after trying to implement the methodology described in Section 3. We also identify the hashtags that come together with the #5G hashtag. The most frequently used are #iot, #technology, and #huawei (see Table 1).

Table 1: Hashtags frequency

Hashtags	Frequency
#iot	1, 196
#technology	364
#huawei	339
#digitaltransformation	287
#tech	265
#cybersecurity	259
#bigdata	255
#wireless	232
#mobile	223
#blockchain	217
#machinelearning	197
#innovation	191
#security	189
#smartcities	174
#telecom	166
#artificialintelligence	163

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

Our classifier got a decent score for positive tweet recognition, but it was lower for neutral and suitable for negative tweets. Overall, the score was satisfactory at 80.89 percent (see Table 2).

Table 2: Sentiment analysis results

Sentiment	Tweets	Precision	Recall
Positive	5,244	84.03%	95%
Negative	906	54.5%	20.05%
Neutral	1,938	73%	69%
Total	8,996	81%	80.89%

Table 2 shows the results of automatically and manually labelling tweets. It's worth noting that the majority of the topics are about tech, economics, and news. Surprisingly, public health is also an issue.

5. DISCUSSIONS

Previous research has shown that social networks are an effective tool for analysing user opinions after the fact or in real time. Twitter, in particular, promotes the interaction of comments through the use of hashtags, which has a number of advantages. One of them is that an individual can express their opinion using the same hashtag, which aids in the generation of thousands of data clusters in a single label "#" (Small, 2011). Currently, one of the most common platforms for extracting data from various sources, many of which are used for analysis, exists. The current research focuses on the sentiment analysis of tweets in which users shared their opinions by using the hashtag #5G in their tweets. Following this approach (see Section 3), we conclude that a sentiment analysis can be performed in an adaptable manner, identifying the text into three categories: positive, negative, and neutral. According to our findings, the statistical sentiment analysis tool obtained 64.35% of 8,102 tweets using Twitter's API as positives, 24.47% as negatives, and 11.17% as neutral (see Table 2). This indicates that the grades are not evenly distributed, since there are more tweets distribute as positives. The findings of this study indicate that #5G technology has a strong presence in the market, with 64.35% of positive

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

tweets. However, the listed topics bring to light some user issues, such as the impact of radiation on people, as well as a cancer-related idea that is part of the conversation.

In addition, we can see in Table 2 that there will be an effect until the 5G network is open to all. With faster networks, emerging technologies such as artificial intelligence, machine learning, and the internet of things, among others, can be used.

6. CONCLUSIONS

As previously mentioned in this paper, Twitter has evolved into a social network that can be used to extract user opinions and ideas on a variety of topics. We review the publication about the 5G Network using the social media platform Twitter in this study. When opposed to the 4G network, this technology is the next level of broadband for mobile communications; it offers a greater coverage, quicker and better times. After cleaning the results, we got a sample of 8,996 tweets. We classify tweets into three groups using statistical sentiment analysis tool: positives, negatives, and neutrals. We can see facts about the benefits, drawbacks, promises, concerns, and fears for the arrival of the 5G network by analysing sentiment. The findings of this study may be useful to technology companies focusing on technological advancements, as they may be able to predict future behavior or positions on a given subject. Companies could develop strategies to deal with user sentiment based on this information.

REFERENCES

Agarwal, A., Agarwal, K., Agarwal, S., & Misra, G. (2019). Evolution of Mobile Communication Technology towards 5G Networks and Challenges. *American Journal of Electrical and Electronic Engineering*, *7*(2), 34–37.

Bermingham, A., & Smeaton, A. F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results. *Psychology*, 2–10.

Cavazos-Rehg, P. A., Zewdie, K., Krauss, M. J., & Sowles, S. J. (2018). “No High Like a Brownie High”: A Content Analysis of Edible Marijuana Tweets. *American Journal of Health Promotion*, *32*(4), 880–886.

Chamlertwat, W., Bhattarakosol, P., Rungkasiri, T., & Haruechaiyasak, C. (2012). Discovering consumer insight from twitter via sentiment analysis. *Journal of Universal Computer*

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

Science, *18*(8), 973–992.

Filho, R. M., Almeida, J. M., & Pappa, G. L. (2015). Twitter population sample bias and its impact on predictive outcomes: A case study on elections. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015*, *April 2017*, 1254–1261.

Global Digital Report: Digital in 2019. (n.d.). Retrieved August 30, 2019, from https://wearesocial.com/globaldigital-%0Areport-2019%0A

Go, A., Bhayani, R., & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *Processing*, 1–6.

Gohil, A., Modi, H., & Patel, S. K. (2013). 5G technology of mobile communication: A survey. *2013 International Conference on Intelligent Systems and Signal Processing, ISSP 2013*, *February 2015*, 288–292.

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online Learning for latent Dirichlet allocation (Supplementary Material). *Nature*, 1–9. http://papers.nips.cc/paper/3902-online-learning-for-latent-dirichlet-allocation.pdf

Holzinger, A. (2017). Introduction to MAchine Learning & Knowledge Extraction (MAKE). *Machine Learning and Knowledge Extraction*, *1*(1), 1–20.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Loyola-Gonzalez, O., Monroy, R., Rodriguez, J., Lopez-Cuevas, A., & Mata-Sanchez, J. I. (2019). Contrast Pattern-Based Classification for Bot Detection on Twitter. *IEEE Access*, *7*, 45800–45817.

Mane, S. B., Sawant, Y., Kazi, S., & Shinde, V. (2014). Real Time Sentiment Analysis of Twitter Data Using Hadoop. *International Journal of Computer Science and Information Technologies*, *5*(3), 3098–3100.

Martin-Domingo, L., Martín, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). *Journal of Air Transport Management*, *78*(January), 106–115.

Neethu, M. S., & Rajasree, R. (2013). Sentiment analysis in twitter using machine learning techniques. *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*.

Reyes-Menendez, A., Saura, J. R., & Alvarez-Alonso, C. (2018). Understanding #worldenvironmentday user opinions in twitter: A topic-based sentiment analysis approach. *International Journal of Environmental Research and Public Health*, *15*(11).

Samuels, A., & McGonical, J. (2020). Sentiment Analysis on Social Media Content. *ArXiv*.

ISSN: 2229-6913(Print), ISSN: 2320-0332(Online)
Received: 10-12-2020; Revised: 15-03-2021; Accepted: 15-05-2021

Saura, J. R., Palos-Sanchez, P., & Martin, M. A. R. (2018). Attitudes expressed in online comments about environmental factors in the tourism sector: An exploratory study. *International Journal of Environmental Research and Public Health*, *15*(3).

Small, T. A. (2011). What the hashtag?: A content analysis of Canadian politics on Twitter. *Information Communication and Society*, *14*(6), 872–895.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, *July*, 115–120.

Wang, W., & Wu, J. (2011). Emotion Recognition Based on CSO&SVM in E- Learning. *2011 Seventh International Conference on Natural Computation*, *1*, 566–570.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, *55*(4), 919–926.