Covid-19 Public Sentiment analysis Using Tweet Classification with Machine learning: Review

Nitika Khanna[1], Shivani Rana[2]
[1, 2] *Department of Computer Science and Engineering*
Himachal Pradesh Technical University, Hamirpur, Himachal Pradesh, India

Abstract: In addition to the Coronavirus outbreak, another severe crisis occurred in the form of mass fear and panic, being part of unreliable and sometimes inaccurate information. A detailed understanding of COVID-19's informational problem and assessing public opinion are essential to consider the issue and allocate resources adequately. We classify public opinion correlated with Pandemic A with Coronavirus related Tweets and R programming tools and its sentiment analysis packages. As a result of Google's algorithms, people will begin to believe that they are not being handled equally online. We study two main machine learning approaches and address the scope of textual analysis. Tweets showed relatively high accuracy with the Naïve Bayes system. Our system is reliable, with over 72% accuracy, but it continues to gain consistency and strength as the Tweets' duration increases. The analysis offers an insight into the sentiment advancement of Coronavirus and highlights consequences, drawbacks, and opportunities.

Keywords: *tweet, covid, analysis, machine learning*

I. INTRODUCTION

In this study, we covered four crucial issues: (A.I.). Coronavirus and COVID-19 infection have deeply stressed the need for data analytics methods to better understand the knowledge flow during a pandemic scenario. While there have been numerous initiatives to analyze healthcare, preventative, care and recovery, economic and network data, this has provided a relatively little account of personal and social media communications. McKinsey described the essential aspects of Global 3C.

For the COVID-19 recession scenario. In their paper, this organization's researchers emphasize that data management, monitoring, and informational dashboards are significant resources to handle COVID-19 scenarios.

There has been a substantial increase in adopting methods for studying written and spoken discourse (textual analytics and natural language processing). Even with NLP methods, there are issues regarding deciphering the intrinsic meaning in a text. The researchers also demonstrated how even the most recent artificial intelligence schemes could leaving a distance, they are also prone to undermining terms. You must be informed about texts classification methods and machine learning algorithms because they can be helpful in the job market. It is important to investigate whether complementary types of artificial intelligence system will enhance the current computer system. Studies on electronic economies reveals the usefulness of machine learning in predicting human actions in dynamic scenarios. There is Twitter data being used in this article.

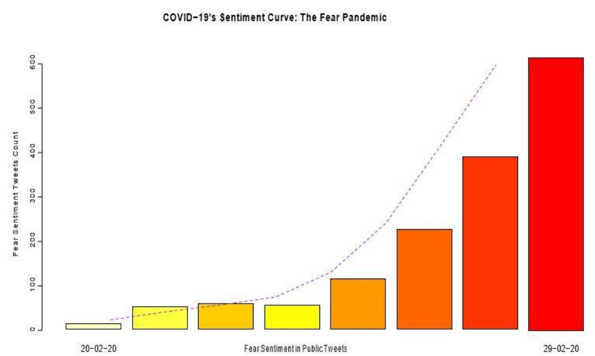

Figure 1. Fear curve.

These patterns are partially due to the general public's dependence on conventional news outlets to gather information, rather than social media. People's attitudes and attitudes on different social phenomena are created because they are linked to new technologies. Many businesses are using social media platforms to gain more attention. This statement captures the nature of social media's direct connections between people. Social networking and an open communication forum have become an essential source of information and ideas for us. Researchers and practitioners mine the textual and unstructured datasets to understand mass activity and emotional conditions through statistical analysis. Motivational values and stock market sentiment are associated [13]. The textual data visualization is also used to describe the pattern of change in anxiety feelings over the period in Figure 1, using the "Lowes curve" and the bars showing the regular growth of anxiety tweets count. As it was the only emotion that was sent through Tweets, the examination of "fear-sentiment" in Tweets. This data revealed that fear-sentiment would increase at the end of March 2020.

This research article presents a report on defining the attitudes of the public. This study proposes an approach to examine twitter data explicitly to recognize sentiment, interaction with keywords, and patterns in crises such as the COVID 19 phenomena. We were able to proceed with the discussion and the quest for insights with descriptive textual analytics such as exploratory Word Clouds and sentiment maps. Generating topic-modeling out of Twitter Tweets in a sequential fashion revealed interest behind certain words, as shown in Table 1. We found that associating beer and humor were overtaken by fear (Figure 1) during this research. Tweets insights aim to examine the text and recognize changes in public opinion.

Table 1. Tweet features summarized by source category.

Source	Total	Hashtags	Mentions	Urls	Pols	Corona	Flu	Beer	AbuseW
iPhone	3281	495	2305	77	218	4238	171	336	111
Android	1180	149	1397	37	125	1050	67	140	41
iPad	75	6	96	4	12	85	4	8	2
Cities	30	0	0	0	0	0	0	0	0

One of this project's main contributions is the debate, presentation, and comparison of Naïve Bayes and Logistic techniques, which are popularly used in A.I. use in Twitter. The consistency is measured by the sum of correct decisions to the total number of used objects. We identified that Naïve Bayes is the best technique for short Tweet sentiment classification and can evaluate short Coronavirus Tweets with an accuracy of 91%. For a longer tweet, naïve Bayes provided 57% accuracy, while logistic regression provided 52% accuracy.

II. LITERATURE REVIEW

this paper deals with researching literature in textual analytics and the machine learning method. Machines help obtain accurate findings from vast quantities of raw data. Text analytics includes analysing characters, syntax, semantics, sentiment, and graphic representations of documents. Exogenous variables include message length, the use of keywords, special characters, connections, and hashtag. A list of popular Twitter symbols used in this particular data point is given.

Table 2. Summary of endogenous features.

Tagged	Frequency	Hashtag	Frequency
realDonaldTrump	74	coronavirus	23
CNN	21	DemDebate	16
ImtiazMadmood	16	corona	8
corona	13	CoronavirusOutbreak	8
AOC	12	CoronaVirusUpdates	7
coronaextrausa	12	coronavirususa	7
POTUS	12	Corona	6
CNN MSNBC	11	COVID19	5

The external variables are introduced, but they are not central. The table summarizes "type of device used" and "screen name," showing external to the message's text and the Twitter user's screen name, respectively. Such summaries help in describing the primary characteristics of the results. Data scientists started to use texts to find new methods for detecting fraud.

Table 3. Summary of exogenous features.

Source	Frequency	Screen Name	Frequency
Twitter for iPhone	3281	_CoronaCA	30
Twitter for Android	1180	MBilalY	25
Twitter for iPad	75	joanna_corona	17
Cities	30	eads_john	13
Tweetbot for i<U+039F>S	29	_jvm2222	11
CareerArc2.0	14	AIAboutNothing	11
Twitter Web Client	16	dallasreese	9
511NY-Tweets	3	CpaCarter	8

a. Textual Analytics

According to nature, size of data, research goals, and context, multiple tools were used for textual analytics. Twitter has been used frequently for emotion analysis [18-20]. Research using Twitter to evaluate consumer reviews for a French Energy Firm uncovered fascinating insights about the company not acknowledged due to the number of documents. Poisson and negative binomial regression models were used to analyze Tweet popularity. We can see that Kullback-Leibler and Euclidean distances work well in deciding similar topics for network analysis. Analysis using TAKE methodology showed methods for discovering useful knowledge from the masses of data available on Facebook and Twitter. This analysis used topic-based summarizing of Twitter data to examine research interests on the Internet. They used methods of the summary to conduct high-quality analysis. Twitter data has been utilized to ascertain personality categories using DISC techniques. Similar strategies have been used concerning information systems regarding recognizing human characteristics, including the superiority of electronic communication. DISC assessment can be used for information processing, product placement, and psychological screening dispositions. Also, there is a psychological and linguistic analysis used in previous studies to derive feelings from social media messages.

b. Twitter Analytics

According to current research, social media data is useful in exposing situational knowledge during crisis situations. The study has examined approximately 41,545 wildfire-related tweets from May of 2014. A comparative analysis of firefighters' and the population data found that six of the nine largest wildfires occurred on May 14. According to Kernel Density Estimate, there were various tweets in San Diego with keywords related to fires and wildfires on Twitter. This example illustrates that there is a separation between reality and tweets. Analysis of Twitter data in the current research also revealed some inconsistencies between Twitter sentiment, and As the population of urban areas are more significant, they have

a higher number of Internet connections, resulting in a higher number of tweets from urban areas. The same research in San Diego showed that people posting about fires were more likely to share their feelings and concerns. Tweets also showed appreciation of California firefighters and the containment of wildfires. Twitter promoted improved situational awareness and increased emergency management efforts in the wildfires.
Twitter has been used for numerous reasons, including handling crises. Nagar et al. [29] validated the predictive power of regular Twitter data for emergency room visits for influenza-like illness during the New York City 2012–2013 influenza season. Widener and Li (2014) [8] analyzed the regional distribution of sentiment in tweets containing the words "healthy food" and "unhealthy food" in the U.S. In general, urban and suburban residents tweet more than those residing in rural areas. Meal tweets were higher per capita in major metropolitan centers than in small towns. Logistic regression showed that tweets with unhealthy foods were more likely to be released in low-income areas. Twitter data has also been used in the sense of security and mental health. In a study by De Choudhury et al. (2013) [11], researchers explored mood of new mothers in postnatal situation. This research used Twitter posts to show that such data is very good at detecting postpartum depression in women. New analytic methodologies and frameworks have been employed to review the Twitter postings made about supply chain management (SCM). They analyzed descriptive analytics, content analysis, and network analytics of 22,399 SCM tweets. Carvalho et al. [31] were the first to use Twitter to perform research and development regarding culture's human influence. This platform is excellent in allowing non-technical people to mine social media contents easily.

2.3 **Classifying approaches.**

Extant research uses different forms of text classification for the sentiment on social media. These classifiers are based on several similarities. The next section reviews the basic principles of two classifiers we tested: naïve Bayes and logistic regression. My study focuses on defining the effective predictive strategies of Naïve Bayes and logistic regression.

2.3.1**. Linear Model**

Regression is commonly used to explain relationships between continuous variables. The linear classifier is very functional because it minimizes an objective function (i.e., squared difference between the predicted outcomes and dedicated classes). The least-squares approach is close to the maximum likelihood approach because measured results are distorted by noise. With MLE Classifier, we can obtain full gain out of any misclassification. Ridge Classifier classifies minimal data as either zero (0) or non-zero (1) and handles multi-class regression problems.

2.3.2. **Neural Network Classifier**

Naïve Bayes classifier is an easy and proven technique for classification. E-mail is confidential paper since the 1950s. The formula in question is based on the Bayes theorem. A mathematical formulation is given. NBC uses maximal a posteriori estimate of the likelihood distributions (i.e., features are assigned to a class based on the highest conditional probability). There are two types of Naïve Bayes models: Bernoulli Naïve Bayes and Multinomial Naïve Bayes. A lot of fascinating projects involving NEBs have been carried out in academic institutions. Researchers have demonstrated that NBC has a better accuracy in detecting documents than classification methods such as decision trees, neural networks, and support vector machines. The Center for Calculating University Success identified more than 77% precision of IEEE for forecasting the influential position of students in the university. Another study indicates that the hitting average for symbolic methodology was under 80%. Machine learning has the potential to identify emotion with over 80 percent accuracy. NBCS is conveniently applicable for minimal data only. Yet, Generative belief networks have basic errors in algorithms.

2.3.3. **Logistic regression analysis**

Logistics regression is one of the traditional methodologies of classification. L.R. was first proposed in 1958. In the logistic regression model (L.R.), the outcome for a single trial is modeled using a logistic function. Then we can use the logistic function to transform the results into binary values (0 and 1). Maximum probability projections are widely used in both models. The Tableau Collaborative claims that the Logistic Regression Multi-Class Classifier has the best (min 32.43 %, max 58.50 %) accuracy against Naïve Bayes, Random Forest, Decision Tree Support Vector Machines Classifier. Multinomial logistic regression (MLR) was proven accurate in forecasting Twitter users' emotions to 79 percent. The stepwise logistic discriminant regression is very successful in classifying 97% of the cases. L.R. classifier is most useful for dichotomous results. Data points need to be independent of each other. While the classifier is exceptionally accurate, the accuracy declines after several weeks. L.R. classifiers use intelligent algorithms.

2.3.4. **K-Nearest Neighbor**

K-Nearest Neighbor (KNN) takes new data and uses instances to classify recent events. KNN Methodology uses text similarity test to group papers. The similarity is calculated by computing the distance between two points. KNN algorithm counts the vote of closest neighbors to the user. If an issue is the nearest neighbor is defined by either description or interpolation. This approach is beneficial, being very easy to execute.

Classifier	Characteristic	Strength	Weakness
Linear regression	Minimize sum of squared differences between predicted and true values	Intuitive, useful and stable, easy to understand	Sensitive to outliers; Ineffective with non-linearity
Logistic regression	Probability of an outcome is based on a logistic function	Transparent and easy to understand; Regularized to avoid over-fitting	Expensive training phase; Assumption of linearity
Naïve Bayes classifier	Based on assumption of independence between predictor variables	Effective with real-world data; Efficient and can deal with dimensionality	Over-simplified assumptions; Limited by data scarcity
K-Nearest Neighbor	Computes classification based on weights of the nearest neighbors, instance based	KNN is easy to implement, efficient with small data, applicable for multi-class problems	Inefficient with big data; Sensitive to data quality; Noisy features degrade the performance

REFERENCES	Motivation	Datasets	Methods	Limitations
[32]	I am using artificial intelligence to predict the growth of panic amongst Twitter users following a particular keyword.	Data was collected using Twitter API. Around 9,000,000 tweets were analyzed in this study.	Naïve Bayes and Logistic Regression classifiers have been used to train the classifiers. The precision for shorter tweets was ninety-one percent (91%) and seventy-four percent (74%), respectively.	This paper examined the conversations and discovered and arranged them on the basis of one keyword. Further dimensions that cannot be analyzed by geographical limits should also be discussed.
[33]	Analyze the thoughts of Indians after the enforced lockdown by the government.	Datasets are gained via Twitter R's programming guide. Approximately 24,000 tweets with "#IndiaLockdown" and "#IndiafightsCorona" within the period of 25th to March 28, 2020, were extracted.	The research was done by using R software and constructed using Word Cloud.	Few tweets from around the world were considered. In the study conducted by IISER, Indians regarded the steps taken by the government positively.
[34]	The Lucc Tracker provides political and health-related effects on the people as envisaged through the site...	The source was taken from the website "Corona Tracker."	Predictive models were used to predict the rate of the disease to spread.	The news reveals that it is from the beginning of March 2020. No data was obtained.
[35]	seize the nature of the primary subject from tweets linked to the COVID-19 epidemic.	1,68,000 tweets were grouped into 13 different subjects.	Twitter tweets were described through unigrams and bigrams for subject modeling.	Though twelve topics of interest were established during the space of February '20 to March '20, it was more among topics of sentiments. These were linked to medical problems.
[36]	Look at how Chinese users of Weibo were influenced emotionally on January 20, 2020.	This is essential data for analyzing the use of Weibo by Weibo users in Jan 2020.	A paired sample t-test by SPSS (Statistical Product and Service Solutions) reveals the Weibo users' emotional elements.	Twitter users are mainly young. The result was not unbiased. This study demonstrates how the negative sentiment develops among the youths after the outbreak of the pandemic.
[37]	Tweets is built on the format of twitter and is an important source of data for academic study.	metadata which consists of title and abstract of the experiment.	Frequent Patterns Mining using an adapted FP-Growth algorithm.	Fix number of pattern reduce accuracy

[38]	The case extractions are pertinent for epidemiologists. By writers or decision-makers, helping. Individuals search some bit of information in order to obtain awareness. Moreover, Our tools may allow new kinds of data aggregations. information analytic.	Tweet API	up-to-date a BERT (Devlin et al.,. 2019 has dependent classifier that processes an English text as input. and encloses a nominee slot in the tweet.	Unsupervised learning increase cost of annonation
[39]	the public events, industry, educations, welfare activities It happens to all in life. Influences. Their jobs are being lost and their welfare is getting worse. Stress is growing on both personal and communal levels. ages. Study in behavioural economics.	Twitter surveys. "API2" and "Tweepy3". Here we have compiled over 700gigabytes of raw data. After receiving data of the volunteers until April 24, 2020, I stored the data as JSON files.	Polarity for the various states is seen in a set of graphs. 2. Figure 2 shows a Google Map you can zoom in and switch about. Sentiment of the condition can be observed here. For a period of time. The term cloud (c) reinforces these predictions. highlight what has cause some people to feel that way (c) Figure 2 indicates the environmental polarity over time.	Only analysis not classified

III. CONCLUSIONS AND FUTURE WORK

We discuss concerns affecting public opinion as fears intensify for avian influenza and Coronavirus spread. We also illustrated the use of exploratory and informative textual analytics and textual data visualization techniques. We found certain items during the text analysis, such as the grouping of words by amounts of a specific non-text variable. We also compared the efficacy of different text classifiers in processing Tweets and showed their importance in data analysis forms. The present research provided approaches with useful knowledge and public opinion insights generation capacity. To solve the global data challenges of COVID-19, there is a need for a comprehensive range of advanced artificial intelligence and computer-based methods. A lot of research and studies need to be conducted on social media, news and the Internet. It is essential to adapt sufficient contact approaches post-COVID-19 to meet explicitly public health needs. Corporations and small enterprises may also benefit from automated machine learning models. Our analysis continues, expanding upon the infrastructure previously built in this report.

IV. REFERENCES

[1]. Online: **https://www.mckinsey.com/~/media/mckinsey/business%20functions/risk/our%20insights/covid%2019%20implications%20for%20business/covid%2019%20may%2013/covid-19-facts-and-insights-may-6.ashx** (accessed on June 11, 2020).

[2]. Jin, D.; Jin, Z.; Zhou, J.T.; Szolovits, P. Is Bert robust? natural language attack on text classification and entailment. *arXiv* , **2019**, arXiv:1907.11932.

[3]. Samuel, J. Information Token Driven Machine Learning for Electronic Markets: Performance Effects in Behavioral Financial Big Data Analytics. *JISTEM J. Inf. Syst. Technol. Manag.* **2017**, *14*, 371–383

[4]. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explore. Newsl.* **2017**, *19*, 22–36

[5]. Makris, C.; Pispirigos, G.; Rizos, I.O. A Distributed Bagging Ensemble Methodology for Community Prediction in Social Networks. *Information* **2020**, *11*, 199.

[6]. Heist, N.; Hertling, S.; Paulheim, H. Language-agnostic relation extraction from abstracts in Wikis. *Information* **2018**, *9*, 75.

[7]. He, W.; Wu, H.; Yan, G.; Akula, V.; Shen, J. A novel social media competitive analytics framework with sentiment benchmarks. *Inf. Manag.* **2015**, *52*, 801–812.

[8]. Widener, M.J.; Li, W. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the U.S. *Appl. Geogr.* **2014**, *54*, 189–197.

[9]. Kretinin, A.; Samuel, J.; Kashyap, R. When the Going Gets Tough, The Tweets Get Going! An Exploratory Analysis of Tweets Sentiments in the Stock Market. *Am. J. Manag.* **2018**, *18*.

[10]. De Choudhury, M.; Counts, S.; Horvitz, E. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In Proceedings of the SIGCHI conference on

human factors in computing systems, Paris, France, April 27 –2 May 2013; pp. 3267–3276

[11]. Wang, Z.; Ye, X.; Tsou, M.H. Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Nat. Hazards* **2016**, *83*, 523–540.

[12]. Skoric, M.M.; Liu, J.; Jaidka, K. Electoral and Public Opinion Forecasts with Social Media Data: A Meta-Analysis. *Information* **2020**, *11*, 187.

[13]. Samuel, J. Eagles & Lions Winning Against Coronavirus! 8 Principles from Winston Churchill for Overcoming COVID-19 & Fear. Researchgate Preprint. 2020. Available online: **https://www.researchgate.net/publication/340610688** (accessed on April 21 2020).

[14]. Chen, X.; Xie, H.; Cheng, G.; Poon, L.K.; Leng, M.; Wang, F.L. Trends and Features of the Applications of Natural Language Processing Techniques for Clinical Trials Text Analysis. *Appl. Sci.* **2020**, *10*, 2157.

[15]. Reyes-Menendez, A.; Saura, J.R.; Alvarez-Alonso, C. Understanding# WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2537.

[16]. Saura, J.R.; Palos-Sanchez, P.; Grilo, A. Detecting indicators for startup business success: Sentiment analysis using text data mining. *Sustainability* **2019**, *11*, 917.

[17]. Samuel, J.; Holowczak, R.; Benbunan-Fich, R.; Levine, I. Automating Discovery of Dominance in Synchronous Computer-Mediated Communication. In Proceedings of the 2014 47th IEEE Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 1804–1812.

[18]. Rocha, G.; Lopes Cardoso, H. Recognizing textual entailment: Challenges in the Portuguese language. *Information* **2018**, *9*, 76.

[19]. Carducci, G.; Rizzo, G.; Monti, D.; Palumbo, E.; Morisio, M. Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information* **2018**, *9*, 127.

[20]. Ahmad, T.; Ramsay, A.; Ahmed, H. Detecting Emotions in English and Arabic Tweets. *Information* **2019**, *10*, 98.

[21]. Pépin, L.; Kuntz, P.; Blanchard, J.; Guillet, F.; Suignard, P. Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets. *Comput. Ind. Eng.* **2017**, *112*, 450–458.

[22]. De Maio, C.; Fenza, G.; Loia, V.; Parente, M. Time aware knowledge extraction for microblog summarization on twitter. *Inf. Fusion* **2016**, *28*, 60–74.

[23]. Ahmad, N.; Siddique, J. Personality assessment using Twitter tweets. *Procedia Comput. Sci.* **2017**, *112*, 1964–1973.

[24]. Jain, V.K.; Kumar, S.; Fernandes, S.L. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *J. Comput. Sci.* **2017**, *21*, 316–326.

[25]. Ye, X.; Li, S.; Yang, X.; Qin, C. Use of social media for the detection and analysis of infectious diseases in China. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 156.

[26]. Fung, I.C.H.; Yin, J.; Pressley, K.D.; Duke, C.H.; Mo, C.; Liang, H.; Fu, K.W.; Tse, Z.T.H.; Hou, S.I. Pedagogical Demonstration of Twitter Data Analysis: A Case Study of World AIDS Day, 2014. *Data* , **2019**, *4*, 84.

[27]. Kim, E.H.J.; Jeong, Y.K.; Kim, Y.; Kang, K.Y.; Song, M. Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *J. Inf. Sci.* **2016**, *42*, 763–781.

[28]. Samuel, J.; Ali, N.; Rahman, M.; Samuel, Y.; Pelaez, A. Feeling Like it is Time to Reopen Now? COVID-19 New Normal Scenarios Based on Reopening Sentiment Analytics. *arXiv* , **2005**, arXiv:2005.10961.

[29]. Nagar, R.; Yuan, Q.; Freifeld, C.C.; Santillana, M.; Nojima, A.; Chunara, R.; Brownstein, J.S. A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J. Med Internet Res.* **2014**, *16*, e236.

[30]. Chae, B.K. Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *Int. J. Prod. Econ.* **2015**, *165*, 247–259.

[31]. Carvalho, J.P.; Rosa, H.; Brogueira, G.; Batista, F. MISNIS: An intelligent platform for twitter topic mining. *Expert Syst. Appl.* **2017**, *89*, 374–388.

[32]. Vijayan, V.K.; Bindu, K.; Parameswaran, L. A comprehensive study of text classification algorithms. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017; pp. 1109–1113.

[33]. Zhang, J.; Yang, Y. Robustness of regularized linear classification methods in text categorization. In Proceedings of the 26th Annual International ACM SIGIR Conference On Research and Development in Informaion Retrieval, Toronto, ON, Canada, July 28 2003; pp. 190–197.

[34]. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

[35]. Liu, B.; Blasch, E.; Chen, Y.; Shen, D.; Chen, G. Scalable sentiment classification for big data analysis using naive bayes classifier. In Proceedings of the 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA, 6–9 October 2013; pp. 99–104.

[36]. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150.

[37]. Drias, H. H., & Drias, Y. (2020). Mining Twitter Data on COVID-19 for Sentiment analysis and frequent patterns Discovery. *medRxiv*.
[38]. Zong, S., Baheti, A., Xu, W., & Ritter, A. (2020). Extracting COVID-19 Events from Twitter. *arXiv preprint arXiv:2006.02567*.
[39]. Kabir, M., & Madria, S. (2020). CoronaVis: A Real-time COVID-19 Tweets Analyzer. *arXiv preprint arXiv:2004.13932*.
[40]. Troussas, C.; Virvou, M.; Espinosa, K.J.; Llaguno, K.; Caro, J. Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. In Proceedings of the IISA 2013, Piraeus, Greece, 10–12 July 2013; pp. 1–6.
[41]. Ting, S.; Ip, W.; Tsang, A.H. Is Naive Bayes a good classifier for document classification. *Int. J. Softw. Eng. Appl.* **2011**, *5*, 37–46.