Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Applications of sentiment analysis and machine learning techniques in disease outbreak prediction – A review

Rameshwer Singh [a,*], Rajeshwar Singh [b]

[a] I.K. Gujral Punjab Technical University Kapurthala, Punjab, India
[b] Doaba Khalsa Trust Group of Institutions SBS Nagar, Punjab, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Machine learning
Outbreak predictions
Twitter data mining
Sentiment analysis
Healthcare

ABSTRACT

Prediction of the facts about any topic always sounds fascinating. It becomes more important and useful when it can predict the facts about healthcare information. In computer science, machine learning techniques/models can be used to predict important health care facts, further the study of the sentiments related to health care information can also play a significant role in the generation of important information. There are many international and national organizations like United States Centers for Disease Control and Prevention (CDC), World Health Organization (WHO), etc. that publish important healthcare data. These data sets can be used to predict important facts about the disease. Further, the datasets related to sentiments of the people like tweets on Twitter, posts on Facebook, and blogs are useful to study the sentiments of the people on a particular topic. Sentiments related to health care information, diseases, and epidemics are very useful. The study of these sentiments can help make better prediction systems and to generate fruitful health care facts. Further, the real datasets related to infectious diseases like COVID-19 can be used for regression analysis for making predictions regarding the upcoming cases. In this paper, we have presented a survey on the application of different machine learning techniques, that are used in the classification and regression analysis related to health care predictions and the study of sentiments. We have presented a review of major research studies related to healthcare between the years 2010 to 2020.

Selection and peer-review under responsibility of the scientific committee of the International Virtual Conference on Sustainable Materials (IVCSM-2k20).

1. Introduction

Microblogging website like Twitter is very beneficial in predicting the hint of upcoming outbreaks. These platforms are closest to the user on which users can share their sentiments. These sentiments can be processed to generate important information related to many domains like health care, elections, reviews, disease, and others Tables 1-3.

Sentiment analysis or opinion mining is referred to as the study of sentiments shared by people on online platforms like Twitter. These kinds of sentiments can be processed using text processing techniques like natural language processing. As per many recent studies, it is proven that machine learning techniques performed better in sentiment analysis and produced better results in the classification of sentiments [1,2,3]. Twitter datasets and official datasets about the epidemic can be used to make prediction systems that can perform better.

As per the study in paper [2,3], machine learning techniques are a better performer in making prediction systems. This paper recommends machine learning techniques for such a kind of study. A study in paper [4], presented both machine learning-based and lexicon approaches. Paper presented step by step phases to perform sentiment analysis.

In the research paper [5], hashtag-based tweets are collected, and then preprocessing is performed on tweets. After performing normalization tweets are processed. As per the study, the POS feature is less useful for performing studies like sentiment analysis or opinion mining.

Machine learning and lexicon-based approaches are compared in the study presented in the paper [6]. The study is conducted based on analyzing the movie review dataset. As per the results, machine learning techniques have performed better than lexicon-based techniques, in terms of providing accuracy.

* Corresponding author.
E-mail address: rameshwer.banga@gmail.com (R. Singh).

https://doi.org/10.1016/j.matpr.2021.04.356
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the International Virtual Conference on Sustainable Materials (IVCSM-2k20).

Table 1
Benefits and drawbacks of different classification techniques.

Machine Learning Techniques/Models	Benefits	Drawbacks
Linear Regression	Simple Implementation Easy Understanding Quick in Training Classification based on features	Only limited to linear applications Unsuitable for many real-life problems The default assumption of input error The assumption of independent features may not be true always.
Logistic Regression	Simple and Fast Consider the role of independent variables on dependent variables. Multiclass classification applicability Convex loss function	Can be applied to linear problems only The extra effort required on the selection of features Noise ratio is high
K-nearest neighbors (KNN)	Easy to understand Less overhead in the adjustment of parameters	The extra care required for the selection of k. The cost of computation is high when working with large datasets.
Decision Tree	Data pre-processing not required No assumption in data distribution Easily understandable results	Possibility of overfitting The tree may turn into a complex shape Loss of information is possible.
Support Vector Machine (SVM)	High accuracy provided It can deal with non-linear data. Fewer chances of overfitting	High computation cost Special care required in the selection of kernel High processing time Less suitable for multiclass problems
Perceptron	Simple and easy Provides good accuracy in classification	Only workable on linear separation problems

Table 2
Use of social media and real datasets to predict outbreaks and epidemics.

Ref .No	Type	Publication	Technique	Data Source	Key points
16	Influenza Epidemic Sensing	ACM, 2010	Linear Regression	CDC and Twitter	0.78 correlation between Twitter and CDC data
18	Use of term H1N1 and Swine Flu	PloS, 2010	Query System Linear Regression	RSS Feeds and Twitter	Increase in use of N1H1 on Twitter from 8.8% to 40.5% between May to December 2009
19	Estimating Flu using Twitter	IEEE 2019	Linear Regression	UAE based Hospital Data and Twitter	0.59 to 0.93 Correlation
20	Sensing Flu with Twitter	Association for Computational Linguistics, 2011.	SVM	Twitter	0.89 correlation
21	Influenza Surveillance by using Twitter data	PloS, 2013	SVM, Logistic Regression & Time Series Analysis	CDC and Twitter.	correlation 0.93
22	Influenza detection based on Twitter	PloS, 2019	Swarm Optimization	CDC, Twitter	IPSO-SVR performed better
23	Swine flu Forecast using Twitter	Springer, Berlin, Heidelberg, 2010.	Time Series	UK surveillance data Twitter Data	High correlation
24	N1H1 in India	Procedia Computer Science, 2015	SVM, Naïve Bayes, Decision Tree Classification, Random Forest and Techniques	Feeds, Twitter	SVM performed better.
30	COVID-19	Asian Pac J Trop Med, 2020	–	Twitter Data	Collected 364,080 tweets representing the COVID-19
31	Influenza forecast	IEEE, 2015	Time Series Model	Twitter Data	
32	COVID-19	ITEE, April 2020	Sentiment Analysis of Tweets	Twitter Data	Tweets Collected Related to COVID-19 Generating significant information to make predictions
33	COVID-19	IJAST (2020)	Model to predict Epidemic outbreak using machine learning techniques	Twitter & Real Data Sets of COVID-19 cases	Predictions and Location data analysis

Table 3
Research Gap Identified.

Year	Ref No	Authors	Disease Covered	Research Gap
2014	[36]	Santos, José Carlos, and Sérgio Matos	Influenza	Only considered Portugal for the proposed work
2015	[27]	Grover, Sangeeta, and Gagangeet Singh Aujla	Swine Flu	Lack of use of Artificial Intelligence Only tweets are used as datasets
2017	[37]	McGough, Sarah F	Zika	Only predicts one parameter for forecasting i.e. confirmed cases
2018	[38]	Nair, Lekha R., Sujala D. Shetty, and Siddhanth D. Shetty	Heart Disease	Not under the category of epidemics It needed to be linked with some health care service provider for working in real-time.
2019	[39]	Maurice, Nduwayezu	Malaria	The study is limited to Nigeria only
2020	[40]	Petropoulos, Fotios, and Spyros Makridakis	COVID-19	Machine learning techniques not used

Three different approaches are presented in the study of paper [7], i.e. machine learning-based, lexicon-based, and hybrid approaches. The outcome shows that for shore sentences lexicon approach may be preferred and machine learning-based models are better performed but preprocessing on data sets is required and large datasets are required for the training purpose. The hybrid approach uses the good features of both models. In this paper machine learning techniques like Naïve Bayes, Maximum Entropy, and SVM are presented.

Similarly, the prediction of box office revenues is presented in paper [8]. In this study data regarding movies is collected from social networking websites. Datasets are processed using a linear regression model and as per the outcome of the study high correlation is found with real datasets. In paper [9], binary classification and three-way classification models are presented. As per the result of the paper three kernel models performed better than others.

In paper [10], the use of K-means and SVM is presented for predictions and classifications. Paper [11], presents a system for predictions regarding US polls of 2012 based on people's sentiments. Pape [12], presents the study of sentiments regarding the public transportation system of Los Angles. In paper [13], a review of various machine learning techniques is presented.

2. Machine learning techniques

Machine learning techniques/models are used to make classification, clustering, and regression analysis. In most of the examples, the system is trained based on training datasets, and after training, the system tries to make an automated classification or can make predictions, etc. The efficiency of the system is tested based on correctly predicted values and test datasets.

Machine learning methods can be generally divided into three categories i.e. Supervised Learning, Unsupervised Learning, & Semi-Supervised Learning.

2.1. Supervised learning

Predefined labeled datasets are used to provide training to the system in the case of supervised learning. The system makes a prediction and classification based on the learning acquired by the system using training datasets [14,15].

2.2. Unsupervised learning

In the case of unsupervised learning, no labeled dataset is provided for the training purpose. The system learns through direct training. Inputs are directly provided to the system and the system automatically identifies important features of datasets and then tries to use learning gained in predictions and classification.

2.3. Semi-Supervised learning

To use the qualities of both supervised and unsupervised learning, semi-supervised learning can be used. In this approach, both labeled and non-labeled datasets are used for the training purpose. It will reduce the overhead of labeling a large amount of data [14,15].

3. Machine learning algorithms

3.1. Linear regression

The linear regression machine learning technique is used to predict the continuous values. It uses the features of input to make predictions and classification. The weighted parameter is used for every training feature [2,3,15].

$$h\theta = \theta 0 + \theta 1\, a1 + \theta 2\, a2 + \ldots \tag{1}$$

Training feature theta is initialized and is updated randomly to minimize the gap between the actual value and predicted value. A gradient decadent algorithm is used to tune the feature in the right direction. Mean squared error is used as the loss function. It can be referred to as the square of the difference between the actual and predicted value.

3.2. Logistic regression

Another classification approach is logistic regression. In this technique outcome is evaluated based on probability (0<=z<=1). It performs binary classification, for example of z < 0.5 then the output will be 0 or 1 otherwise.

In the case of logistic regression Sigmoid function is used. Here Z is the same as used in the case of linear regression. It works based on probability. The value of Z will be 0 when the value of g(Z) is 0.5 and when g(Z) is greater than 0.5 the output will be 1 [2,3,15].

$$Z = \theta 0 + \theta 1 a1 + \theta 2 a2 + \tag{2}$$

$$h(\theta) = g(Z) \tag{3}$$

$$g(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

In this nonlinear Sigmoid function is used and cross-entropy is considered as the loss function. There are two equations in this function one for X = 1 and the other for X = 0 as shown below:

$$J(\theta) = \frac{1}{m} \sum \text{cost}(X', X) \tag{5}$$

$$cost(X', X) = -\log(1 - X') \; if \; X = 0 \tag{6}$$

$$cost(X', X) = -\log(X') \; if \; X = 1 \tag{7}$$

3.3. K-Nearest Neighbor

K-nearest neighbor is used for both classification and regression. It uses the principle of local approximation. It is one of the easiest techniques to understand. It explores its neighbors and analyzes them, to predict the value. In this technique, the lazy learner model is used. In this technique, no loss function is used and no training is involved. It measures the distance between the neighbor and selects the neighbor with minimum distance while doing classification [2,3,15].

3.4. Decision tree

In the case of decision trees, a hierarchical structure is followed. It can be used to solve the problems of classification and regression. It works based on tree structure where each node is labeled with a condition. The condition of each node is tested before marking the node as negative. This process continues till it reaches the leaf node and output is predicted. It works based on independent variables. It uses Gini-index for classification.

$$\text{gini} - \text{index (gi)} = 1 - \sum Z_t^2 \tag{8}$$

Attributes having maximum Gini index value will be selected for the next iteration.

3.5. Naive Bayes

Naïve Bayes works based on the feature matrix. These features are used to calculate the probabilities. It is a supervised learning technique. It is suitable for problems related to classification. There are different methods available in these techniques that can be used to calculate probabilities [7,8,20]. Two popular variants of Naïve Bayes are Multinomial Naïve Bayes, which is used for text classification and second is Bernoulli Naive Bayes, it supports multiple features that are having binary values.

3.6. Support Vector Machine

Support Vector Machine (SVM) is one of the most popular machine learning models. SVM is suited for both classification and regression both. It has different tuning parameters like kernel, margin, gamma, and regularization. Kernel decides which linear function to be used and regularization indicates misclassification. The role of each training dataset is described by gamma and margin defines the margin between the linear separator and points. One of the famous python-based libraries for machine learning techniques is SciKit Learn. It includes two methods SVR for SVM-based regression and SVC for SVM-based classifier.

3.7. Perceptron

Another popular classification technique is perceptron. It is a linear classifier. It performs self-update based on errors and tries to obtain better results. Perceptron is a simple but effective technique. x

Some other famous techniques or variants of algorithms are Passive Aggressive Classifier, Ridge Classifier, and Ada Boost Classifier. A passive-aggressive classifier is very much similar to perceptron except for its tuning parameters. Ridge classifier reshapes the output to −1 and 1 and then classifies the data. Ada Boost Classifier makes multiple copies of datasets and tries to adjust itself based on the mistake in classification.

4. Literature review

The research paper [16], has proposed several methods to identify tweets related to influenza. Proceeding further, the paper also shows its relation with CDC data with the help of various regression models. The paper presents that by using simple linear regression and multiple linear regression-based model 0.78 correlation between the CDC data and tweets can be found.

In the article [17], the authors have proposed a system that has been used to search tweets on Twitter based on influenza-like illness. The study concludes a high correlation between the tweets and the data collected from the CDC. A regression model, that predicts the level of influenza-like illness among the population has been proposed by the system. The upshot of the study is that with the help of Twitter data better results can be provided.

In the research article [18], the usefulness of sentiment analysis has been presented, it has spread awareness regarding health care details among the population. In this article, a study of keywords N1H1 and swine flu has been executed and the results show that tweets including keywords N1H1 have great importance.

Article [19], forecasts influenza early to lessen its impact. A system named tweetluenza has been proposed by the authors based on tweets associated with influenza in the UAE. The linear regression model has been used to predict hospital visits in the future based on the tweets. After that data predictions and data of government health care departments are compared and find the high connection between them.

In paper [20], as a system i.e. beneficial to grab tweets associated with influenza has been proposed. In this paper, the Support Vector Machine is used to perform experiments. The article exhibits a high correlation between Twitter and real data. The outcome of the study shows that there is a 0.89 correlation and 42% of tweets are negative.

Paper [21], presents an identification system that is useful to detect influenza. The results of the system have been compared with the health data of New York and CDC data. SVM, Logistic regression has been implemented in the paper. The proposed system achieved 85% accuracy in the detection of influenza.

In paper [22], the influenza-related data on CDC(US) and Twitter are the roots of the proposed system. influenza-like illness has been used and the survey is based on SVM. The outcome of the system states the proposed system works better with SVM.

Article [23], presents twitter as a handy tool to predict the outbreak of N1H1 during the year 2009. The paper also shows that outcome may be affected if the cleaning of noisy data is not taken care of. Hence with proper analysis of Twitter data, it is beneficial to produce useful health care information. In paper [24], a time-series-based classification of N1H1 data is presented. Different datasets like a tweet from Twitter, data of news agencies to predict the N1H1 outbreak are used in the paper. In this paper SVM, Naïve Bayes, Random Forest, and Decision tree classification techniques are used. as per the outcome, SVM executed better than other techniques.

In the paper [25], the authors present the usage of Twitter data to predict the flu. In the article, the C5.0 algorithm is used and the study is carried out on Philippine tweets. The outcome reveals that C5.0 attain better results as compared to Naïve Bayes. The accuracy achieved is 66%.

Similarly, some more articles present the use of Twitter data and machine learning techniques to predict epidemic outbreaks.

Paper [26], proposes a dengue-based survival system that is based on Spatio-temporal analysis. Similarly, the article [27] has proposed one algorithm that gives the clue about the N1H1 flu. Paper [28] shows the importance of geotagging and machine learning techniques. To forecast the diseases spreading due to mosquitos. Hence all aforementioned reviews reveal that data twitter-based sentiment analysis is beneficial for the predictions of the outbreak and other productive health care information.

The article [29], predicts a system that will be informative to monitor the outbreak of flu. The research has used the data of the year 2016 – 17 and 2017 – 18. Research is confined to Italy only. The study results that Google search queries and Twitter are beneficial for the prediction of the outbreak.

In paper [30], 364,080 tweets were collected with hashtags like #wuhan, #china, #ncov, #2019- nCoV, #corona virus china, #coronavirus outbreak etc. The outcome of the paper claimed that most of the tweets related to coronavirus were from the US only.

In paper [31], the demand of the ICUs for coronavirus patients in Italy is presented and further, it shows the idea of predicting the demand for ICUs through the news on channels, etc.

Paper [32], represents the significance of sentiment analysis on the tweet data sets related to COVID-19. In paper [33], the authors proposed a model that is based on machine learning techniques to perform analysis of Twitter data (sentiment analysis and location information analysis). Further regression analysis is also done on real datasets of COVID-19 to make predictions regarding cases of COVID-19 in the future. Similarly paper [34], presents a comparative analysis of machine learning techniques that can be used to predict the epidemic cases and the classification of tweet datasets.

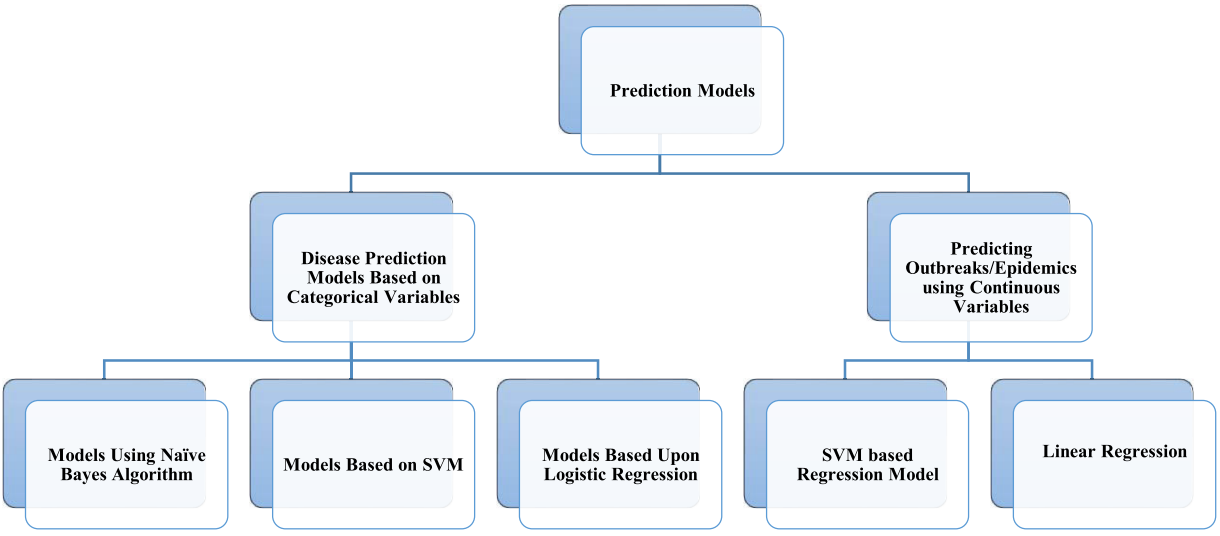

Fig. 1. Types of prediction models.

5. Disease outbreak prediction models

Researchers proposed different models for the prediction of disease outbreaks and their patterns using a Twitter-like microblogging website. Fig. 1 is showing such kinds of models.

In the case of models based on categorical variables, the system is trained to predict the disease based on the input values like health-related featured or symptoms of the patients. Training datasets of such features are used to provide training to the system and then using algorithms like Naïve Bayes, SVM, Logistic Regression based on machine learning techniques [35]. When the predictions of the disease are based on the continuous variable. It is called regression analysis. In such cases, models based on linear regression and SVM are used [35].

As per the review of the various related study, we have identified the following research gap and we will try to overcome the limitation and research gap with the help of our proposed model.

6. Conclusion

Epidemics have a long history and impacted human life many times in history. Many examples show how badly these epidemics have impacted the entire world. The most recent example epidemic is COVID-19 which is affecting the entire globe. COVID-19 is not the only epidemic that is affecting the world but there are many epidemics like Swine Flu, Odisha Jaundice, MERS, Ebola, and many more which have the effect of making an impact on human lives across the world. There is an urgent need to deal with the problems of epidemics with the help of technology.

In this paper, we have gone through many important research studies conducted between 2010 and 2020 in the area of disease prediction. As per the review done in the paper, it is clear that machine learning techniques are better than other techniques to do sentiment analysis. Location information in the tweet carries important information to predict the spread patterns of the disease outbreak. Further, the regression analysis of the real data sets also important to make predictions of disease cases.

Technology is a powerful tool that can be used to design such systems that can predict the epidemic outbreak and its patterns. Many of the studies show that sentiment analysis and analysis of real datasets can help in predicting the early hint and its patterns. In the future, certain models and systems can be proposed to make predictions regarding disease or epidemic outbreaks. Such systems will help in the early detection of the disease and its spread patterns.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ranjan, Sandeep, and Sumesh Sood. "Social network investor sentiments for predicting stock price trends." International Journal of Scientific Research and Review 7.02 (2019): 90-97. Article (CrossRef Link)
[2] Singh, Rameshwer, Rajeshwar Singh, and Ajay Bhatia. "Sentiment analysis using machine learning techniques to predict outbreaks and epidemics." Article (CrossRef Link)
[3] Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing,Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. IEEE, 2013. Article (CrossRef Link)
[4] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal 5.4, Elsevier (2014): 1093-1113. Article (CrossRef Link)
[5] Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!." ICWSM 11 (2011): 538-541. Article (CrossRef Link)
[6] Annett, Michelle, and Grzegorz Kondrak. "A comparison of sentiment analysis techniques: Polarizing movie blogs." Advances in artificial intelligence. Springer Berlin Heidelberg, 2008. 25-35. Article (CrossRef Link)
[7] Kharche, Ms Swapna R., and Lokesh Bijole. "Review on Sentiment Analysis of Twitter Data." International Journal Of Computer Science And Applications 8.2 (2015). Article (CrossRef Link)
[8] Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010. Article (CrossRef Link)
[9] Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011. Article (CrossRef Link)
[10] Li, Nan, and Desheng Dash Wu. "Using text mining and sentiment analysis for online forums hotspot detection and forecast." Decision Support Systems 48.2 (2010): 354-368. Article (CrossRef Link)
[11] Wang, Hao, et al. "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle." Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, 2012. Article (CrossRef Link)
[12] T.TB. Luong, D. Houston, Public opinions of light rail service in Los Angeles, an analysis using Twitter data, Article (CrossRef Link), 2015.
[13] Blum, Avrim L., and Pat Langley. "Selection of relevant features and examples in machine learning." Artificial intelligence 97.1 (1997): 245-271. Article (CrossRef Link)

[14] E. Aydoğan, M. Ali Akcayol, A comprehensive survey for sentiment analysis tasks using machine learning techniques, Article (CrossRef Link), 2016.
[15] Jason Brownlee, Time Series Forecasting as Supervised Learning [online]. Available [as accessed on 31/03/2020] Web Link(Webref Link)
[16] Culotta, Aron. "Towards detecting influenza epidemics by analyzing Twitter messages." Proceedings of the first workshop on social media analytics. ACM, 2010. Article (CrossRef Link)
[17] Achrekar, Harshavardhan, et al. "Predicting flu trends using twitter data." Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on. IEEE, 2011. Article (CrossRef Link)
[18] Chew, Cynthia, and Gunther Eysenbach. "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak." PloS one 5.11 (2010): e14118. Article (CrossRef Link)
[19] Alkouz, Balsam, Zaher Al Aghbari, and Jemal Hussien Abawajy. "Tweetluenza: Predicting flu trends from twitter data." Big Data Mining and Analytics 2.4 (2019): 248-273. Article (CrossRef Link)
[20] Aramaki, Eiji, Sachiko Maskawa, and Mizuki Morita. "Twitter catches the flu: detecting influenza epidemics using Twitter." Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2011. Article (CrossRef Link)
[21] Broniatowski, David A., Michael J. Paul, and Mark Dredze. "National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic." PloS one 8.12 (2013): e83672. Article (CrossRef Link)
[22] Xue, Hongxin, et al. "Regional level influenza study based on Twitter and machine learning method." PloS one 14.4 (2019): e0215600. Article (CrossRef Link)
[23] Szomszor, Martin, Patty Kostkova, and Ed De Quincey. "# swineflu: Twitter predicts swine flu outbreak in 2009." Electronic Healthcare. Springer Berlin Heidelberg, 2012. 18-26. Article (CrossRef Link)
[24] Jain, Vinay Kumar, and Shishir Kumar. "An Effective Approach to Track Levels of Influenza-A (H1n1) Pandemic in India Using Twitter." Procedia Computer Science 70 (2015): 801-807. Article (CrossRef Link)
[25] L.Z. Albances et al., Application of C5. 0 Algorithm to Flu Prediction Using Twitter Data, Article (CrossRef Link), 2018.
[26] J. Gomide et al., Dengue surveillance based on a computational model of spatio-temporal locality of Twitter, Article (CrossRef Link), 2011.
[27] S. Grover, G.S. Aujla, Twitter data based prediction model for influenza epidemic, Article (CrossRef Link), 2015.
[28] Jain, Vinay Kumar, and Shishir Kumar. "Effective surveillance and predictive mapping of mosquito-borne diseases using social media." Journal of Computational Science (2017). Article (CrossRef Link)
[29] C. Comito, A. Forestiero, C. Pizzuti, Improving Influenza Forecasting with Web-Based Social Data, Article (CrossRef Link), 2018.
[30] K. Jahanbin, V. Rahmanian, Using twitter and web news mining to predict COVID-19 outbreak, Article (CrossRef Link), 2020.
[31] Grasselli, Giacomo, Antonio Pesenti, and Maurizio Cecconi. "Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response." JAMA (2020). Article (CrossRef Link)
[32] Rameshwer Singh, Dr. Rajeshwar Singh. "Machine Learning Based Twitter Data Mining to Analyse Sentiments of Tweets Allied to COVID-19 Epidemic & Its Patterns" International Journal of Information Technology and Electrical Engineering Vol. 9, No. 2, April 2020. Article (CrossRef Link)
[33] Rameshwer Singh, Dr. Rajeshwar Singh. "Machine Learning Based Model to Predict Epidemic Outbreak and Pattern Using Official COVID-19 Datasets and Tweets Datasets". International Journal of Advanced Science and Technology, 29(12s), 1808-1821. (2020) Article (CrossRef Link)
[34] Rameshwer Singh, Dr. Rajeshwar Singh. "Comparative Analysis of Machine Learning-Based Techniques/Models for Classification and Regression Using Epidemic Datasets" International Journal of Information Technology and Electrical Engineering Vol. 9, No. 5 October 2020. Article (CrossRef Link)
[35] Leopord, Hakizimana, W. Kipruto Cheruiyot, and Stephen Kimani. "A survey and analysis on classification and regression data mining techniques for diseases outbreak prediction in datasets." Int. J. Eng. Sci 5.9 (2016): 1-11. Article (CrossRef Link)
[36] Santos, José Carlos, and Sérgio Matos. "Analysing Twitter and web queries for flu trend prediction." Theoretical Biology and Medical Modelling 11.S1 (2014): S6. Article (CrossRef Link)
[37] McGough, Sarah F., et al. "Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data." PLoS neglected tropical diseases 11.1 (2017): e0005295. Article (CrossRef Link)
[38] Nair, Lekha R., Sujala D. Shetty, and Siddhanth D. Shetty. "Applying spark based machine learning model on streaming big data for health status prediction." Computers & Electrical Engineering 65 (2018): 393-399. Article (CrossRef Link)
[39] Maurice, Nduwayezu, et al. "Malaria Epidemic Prediction Model by Using Twitter Data and Precipitation Volume in Nigeria." (2019): 588-600. Article (CrossRef Link)
[40] Petropoulos, Fotios, and Spyros Makridakis. "Forecasting the novel coronavirus COVID-19." PloS one 15.3 (2020): e0231236. Article (CrossRef Link)