

Sentiment Analysis and Trend Detection of Tweets using Machine Learning Techniques

Najlaa Musaad Alsadan
Department of Information System
College of Computer and Information Sciences
King Saud University
Riyadh

ABSTRACT

Social media has become the most important source for decision-making procedures in many areas due to its wide applications. People all over the world use their accounts to express personal views, experiences and opinions on diverse topics. Tweets on Twitter are mainly based on the public opinion on a product, event or topic and thus hold large volumes of unprocessed data. Researchers classify and cluster twitter data for different purposes such as Sentiment Analysis, topic detection, topic tracking, culture propagation, opinion mining and others. Analysis of this data is important and difficult due to the size of the dataset. In this review article, we cover most of research articles in tweets classification and clustering. We also address machine learning algorithms for analyzing the data.

Keywords

Sentiment Analysis, Trend Detection, tweets classification, tweets clustering.

1. INTRODUCTION

With the development of Internet technology, social media sites have grown in popularity and played a key role in reconstructing people life [1]. Social media provides people with news and information and makes it possible to exchange feelings with many people. In addition, digital media content can be updated quickly and effortlessly [2].

We will focus on Twitter for several reasons. Twitter is now the primary platform for disseminating information in cases such as natural calamities, conflicts, or events that news reporters cannot cover [3]. Additionally, Twitter, a popular tweeting network, is a very influential place where individuals can express their thoughts on a variety of subjects, often articulating strong opinions in support of or against a particular subject. Twitter is a service that allows people to communicate and stay in touch by sending brief, quick messages to one another. Good. Twitter's design is kept incredibly simple, and the site is easy to use on both mobile devices and computers [4].

People on twitter create new content that becomes the source for some decision-making procedures in many areas. That is because the user data is an important asset that helps us understand the sentiment of the public. We can identify problems of one group of people by investigating another group of people [5]

There are many thousands of Tweets daily. Every user has his favorite subjects to share or retweet. Tweets may be about culture, sports, jobs, education, etc. A lot of researches were done on analyzing twitter data for varied reasons. Some of these reasons are sentiment analysis, topic detection, topic tracking, etc. Business intelligence, and assist in strategic decision making, and can be used in various fields, such as analyzing customer satisfaction, all of that is finally related to sentiment analysis/opinion mining.

We can define topic detection as the topic that has the most daily production of large volumes of messages, Topic tracking or tracking trends is defined as the volume of the daily production of messages on one topic or related hashtag.

We want to carry out a scoping survey to review the fields of tweets analyzing and to find the best machine learning techniques for classifying Twitter data and to present a new survey holding more recent studies. Hence, our research question is: 'What do we know from the existing studies about the use of machine learning techniques in classifying Twitter data? We are particularly interested in the sources of Twitter data; in the data analysis and classification techniques being applied; the accuracy of every technique.

This study has three main goals: covering most of research articles in tweets classification and clustering, address machine learning algorithms for analyzing the data

The rest of the paper has the following sections: review works on Sentiment Analysis of Twitter Data in section 2, review works on Topic Detection and Tracking Techniques on Twitter.in section 3.twitter data collection represented in section 4. Twitter data preprocessing steps represented in section 5. Section 6 represent discussions, finally section 7 represent conclusion and future works.

2. SENTIMENT ANALYSIS OF TWITTER DATA

Sentiment analysis, which is known as opinion mining, is the process of extracting people's opinions, emotions, attitudes, and feelings about a topic or situation from a large amount of unstructured data. Sentiment analysis demonstrates the use of computational linguistics, text mining, Natural Language Processing (NLP) to identify and assess emotional states [6], [7].

In 2022, positive, negative, and neutral emotions in people are examined and categorized. Ensemble techniques, deep learning algorithms, and the machine learning algorithms [8]. Sentiments were classified into positive and negative classes using the Machine Learning (Naïve Bayes, Support Vector Machine, Logistic regression) approaches [9]. The Bi-LSTM approach has achieved better accuracy (0.87) compared to the traditional machine learning models for Twitter sentiment classification. However, the proposed model was evaluated on one dataset only, which makes the worth of the other models under doubt.

Also, Chetanpal Singh et al. [10], present a deep learning approach for sentiment classification of the twitter data related to COVID-19. Their suggested technique uses enhanced featured weighting by attention layers and is based on an LSTM-RNN network. They discovered that the suggested deep learning methodology outperforms competing methods such as Naive Bayes, Random Forest, Support Vector Machine, Logistic

Regression, and LSTM-RNN in terms of accuracy (10%) and precision (10–12%).

Authors in reference [11] proposed a new sentiment classification method called a Hybrid Lexicon-Naïve Bayesian Classifier (HL-NBC). They classified the tweets and removed any irrelevant tweets according to topic classification. In addition, they compared the features of unigram and bigram using the proposed method and Lexicon, Naïve Bayesian classifier. The proposed method has more accuracy of 82% better than other classifiers. In addition, they applied the (HL-NBC) METHOD in an abbreviated time that is less than other methods and has superior results 93% when dealing with big datasets.

In 2021 authors in [7], investigated machine learning approaches along with text processing tools for understanding individuals' sentiments about e-learning during the lockdown following COVID-19. Different machine learning and deep learning models are compared for performance. The outcomes show that when employed with Bow features, the random forest and SVM (Support Vector Machine) classifiers have the maximum accuracy of 0.95 [7]. In 2020, Yadav et al. [12] showed various techniques to analyze the behavior of the tweets as positive, negative or neutral to evaluate and satisfy the sentiments.

Four classification algorithms are considered in that work: the Linear Regression Model, the Naive Bayes Classifier, the Logistic Regression, and the K-Nearest Neighbor. It is one of the works in the field of sentiment analysis related to the pandemic

using the Coronavirus, and it is presented in [13]. Here, the tweet's length is also considered when assessing how well each strategy performs in both short and lengthy tweets. Another method for identifying the crucial trend of change in fear-sentiment is textual data visualization. Four classification algorithms are considered in that work: Linear Regression Model, Naïve Bayes Classifier, Logistic Regression and K-Nearest Neighbor. Here, the length of the tweet is also taken into account when assessing how well each strategy performs in both short and lengthy tweets. Another method for identifying the crucial trend of change in fear-sentiment is textual data visualization.

In [14], they analyzed the effect of two features, TF-IDF word-level and n-gram, on the SS tweet dataset for sentiment analysis. They found that sentiment analysis performed 3-4% better using the TF-IDF word level than using the N-gram function. The analysis is performed using six classification algorithms (Decision Tree, Support Vector Machine, K Nearest Neighbors, Random Forest, Logistic Regression, and Naive Bayes).

According to [15], they proposed a neural network model that covered user behavioral information within his tweets. They used Convolutional Neural Network. They evaluated the system by applying it on two sets of data provided by the SemEval-2016 Workshop. The proposed model used Naive Bayes and Support Vector Machines algorithms. They proved the advantage of going beyond the tweet's contents for sentiment classification.

Table 1. Twitter Sentiment Analysis related studies' survey

Ref. , year	Machine Learning Method	Dataset	Results	Note
[15], 2019	Convolutional Neural Network (CNN)	The SemEval-2016 Stance Dataset	-	
[14], 2019	Logistic Regression and K-Nearest Neighbor, Support vector machines, Random Forest, Decision Tree, and Naive Bayes.	Sentiment Strength Twitter Dataset SS-Tweets dataset	-	
[13], 2020	Naïve Bayes, logistic regression	Coronavirus Tweets of varying lengths	Acc= 91% Acc= 74%	
[11], 2022	Lexicon-Naive Bayesian Classifier (HL-NBC)	Real time Twitter message	82%	
[10], 2022	RNN-LSTM with attention mechanism	COVID-19 Twitter dataset	Acc = 86.12 Pre = 84.23 Rec = 85.23 F1 = 85.12	
[9], 2022	Naïve Bayes, logistic regression, Support vector machines, biLSTM	COVID-19 Twitter Dataset	Acc = 87	there is a need to enhance classification accuracy while keeping computational complexity to a minimum
[8], 2022	Support vector machines, Naïve Bayes, logistic regression, Decision tree, random forest, KNN, Random +logistic, logistic+decision, baseline model, regularized model, dropout model	COVID-19 Twitter dataset	-	

3. TOPIC DETECTION AND TRACKING TECHNIQUES ON TWITTER

As we saw, there are a lot of studies for sentiment analysis of COVID-19 Tweets classification such as [7], [8], [9], [10], and [13]. In addition to sentiment analysis, authors in reference [7] study investigated the effectiveness of e-learning by analyzing the sentiments of people about e-learning. They used topic modelling to identify the issues with e-learning, and the results suggest that the top three issues are the uncertainty surrounding

campus opening dates, children's difficulties understanding online learning, and lacking effective networks for online learning.

Authors in reference [16] proposed a prediction approach of the cryptocurrency market movements of Bitcoin, Ethereum, Ripple, and Litecoin. They used popular machine learning tools and Twitter data. They compared the usage of neural networks (NN), support vector machines (SVM), and random forests (RF) using Twitter elements and market data as input features. The results

demonstrated that cryptocurrency market predictions utilizing machine learning and sentiment analysis are possible.

Governments and healthcare providers may be better able to manage the effects of such diseases on society if they can identify and monitor health conditions like COVID-19 on Twitter. For those businesses, compiling and analyzing tweets on this subject might produce priceless information. From a medical standpoint, sifting through tweets on COVID-19 as a pandemic issue may aid in identifying a cure. Due to the unreasonably high cost of manually processing this information, automatic or semi-automatic ways [17].

One such subject with various applications and significant potential is the automatic detection of traffic occurrences using Twitter data mining. It has been proposed to automatically detect traffic events from Arabic tweets in the Saudi dialect using machine learning and big data technology [18]. Authors build and train a classifier using three machine learning algorithms, Naïve Bayes, Support Vector Machine, and logistic regression, to filter tweets into relevant and irrelevant.

Most of topic detection used clustering techniques. As example reference [19] and [20]. A trending topic detection framework was proposed in 2020 using a method that combines the Transformers with text summarization in a smart way. The sentence embeddings were extracted using Transformer. Then

these embeddings are fed into k-means clustering algorithm to group similar Tweets, and finally text summarization is applied to all sentences of each cluster to supply a summary. The framework was applied to detect main topics on Twitter about the COVID-19 pandemic. Results showed superiority of the proposed approach to other baselines, including Term Frequency-Inverse Document Frequency TF-IDF, and latent Dirichlet allocation (LDA) [19]. Also, reference [21] proposed approach detects the trending topics of the real-time Twitter trends along with ranking the top terms and hashtags. The paper further discusses the motivation for trend prediction over the social media; In addition to exploratory data analysis, the paper explores the Term Frequency-Inverse Document Frequency (TF-IDF), Combined Component Approach (CCA) and Biterm Topic Model (BTM) approaches for finding the topics and terms within given topics.

Event detection, topic extraction, topic tracking, trending topic or Topic detection, as a technique for discovering the main topics automatically, used in many applications that analyze Twitter. There exists a lot of research focused on topic detection on Twitter. About 1692 related articles obtained by a simple search of the Scopus database with the keywords topic + detection + Twitter between 2009 to 2022 [20]. Universal sentence encoder was used as a detection method applied on covid-19 dataset to detect covid-19 topics [20].

Table 2. Twitter topic detection related studies' survey

Reference, year	Machine Learning Method	Dataset	Results	Note
[16], 2019	Multi-layer perceptron	13,096,598 of Bitcoin tweets	Acc.=72%, Pr.= 76%	
	Support vector machines		Acc.=55%	
	random forest		Acc.=61%	
[19], 2020	Universal sentence encoder	COVID-19 dataset	-	COVID-19 topics detection , clustering
[21], 2021	Term Frequency-Inverse Document Frequency (TF-IDF) Combined Component Approach (CCA) Biterm Topic Model (BTM)	Twitter API(63,538 collected tweets)	-	Topic detection, trend ranking , clustering
[20], 2022	sequential k-means (KMEANS) latent Dirichlet allocation (LDA). latent semantic indexing (LSI) non-negative matrix factorization (NFM)	PHEME dataset	Recall= .63% Recall =.71% Recall= .69% Recall= .76%	Clustering

4. TWITTER DATA COLLECTION

For many of studies, Python programming language was used to collect the tweets from Twitter using the library Tweepy as in [22],[23] , and [24]. In addition, some common libraries of python such as NumPy, pandas, and scikit-learn were used with machine learning techniques as in [25],[26] , and [27]. In[28], they used Spritzer to collect the tweets. Spritzer is a sample Twitter Streaming API and returns around 1% of the full Twitter Stream. In [29], they used the Twitter streaming API to collect the data from twitter. The stream API is a Java application that has a connection with a Twitter stream and can store the raw data in a MySQL database.

5. DATA PREPROCESSING

A tweet has many perspectives about the data that are presented by various people in several ways. Social applications offer tools for user interaction or explicit event referencing. Popular social features among Twitter users include hashtags, mentions,

responses, and retweets. A mention is typically used to start a conversation with another user or to include another user in a debate that is already taking place about a specific subject. A post can be replied to or shared again by other users. A hashtag is a special phrase used to find a tweet that begins with "#." A hashtag in a tweet typically alludes to a certain topic, place, or activity. According to the purpose of classification, the preprocessing of data is done. There are specific steps that the developers take [30],[31] , [32] , and [33].

As shown in fig. 1, the basic preprocessing steps are:

- Removing URL and Hashtags. The researchers replaced them with a space string.
- Removing stop words as (In, on, the, and at), it helps in reducing the size of the dataset and the training time. In addition, keep the requested data only.

- Removing punctuation.
- Removing emotions and symbols.
- Stemming: By dropping the suffix, each word was reduced to its base.
- Tokenization: Tokenization is a process that divides lengthy text strings into tokens, which are smaller units of text.

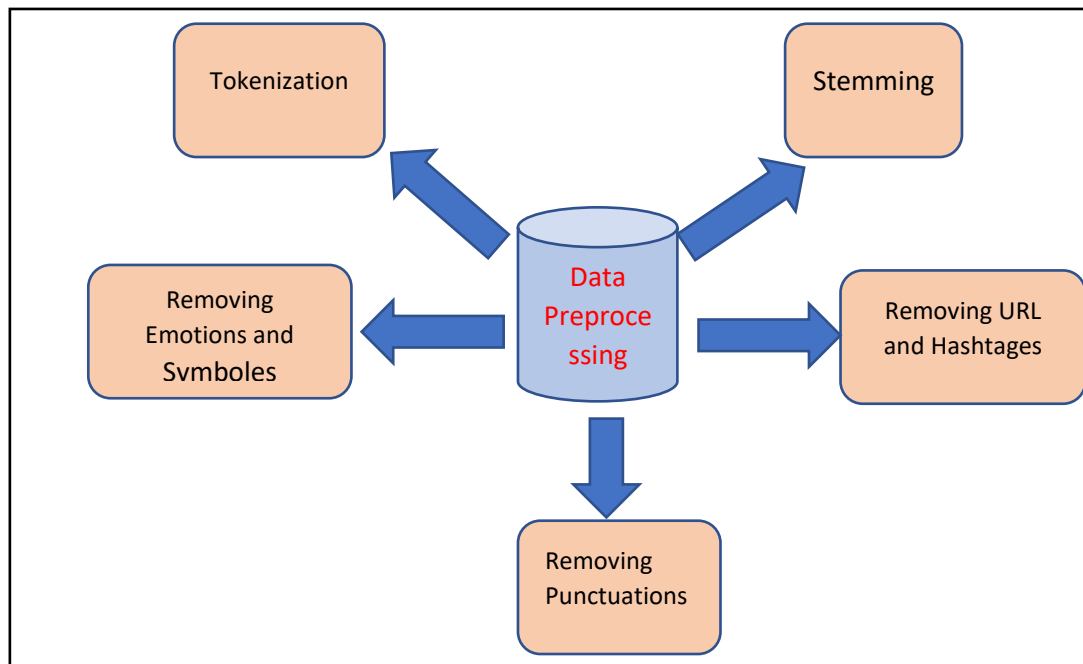


Fig 1: Data preprocessing steps

6. DISCUSSIONS

The key topics that we showed were representative of review of twitter content analysis and detection. Table 1 shows Twitter Sentiment Analysis related studies' survey. We can say that sentiment analysis research has put special focus on Twitter. Table 2 shows Twitter topic and events detection related studies' survey. In the two tables we addressed the most of techniques has been used over the years 2019 until now, we found that tweets analysis based on:

a. Tweets Classification techniques:

The most used machine learning in sentiment analysis techniques is Support vector machines, Naïve Bayes, logistic regression and Decision tree, random forest, Multi-layer perceptron, Support vector machine, K-Nearest Neighbor.

The most used deep learning in sentiment analysis techniques are: Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN), long short-term memory (LSTM), bilstm and dropout model.

b. Tweets clustering techniques:

Clustering algorithms are used to cluster tweet's sentences embeddings. Universal sentence encoder is used to derive the semantic representation and the similarity of tweets [34]. Different clustering algorithms can be used for Tweets Clustering:

- k-means, Sequential k-means (KMEANS) [35]
- Latent Dirichlet allocation (LDA) [36]
- Latent semantic indexing (LSI) [37]
- Non-negative matrix factorization (NMF) [38]

- Term Frequency-Inverse Document Frequency (TF-IDF)
- Combined Component Approach (CCA)
- Biterm Topic Model (BTM)

7. CONCLUSION AND FUTURE WORKS

In this work, we review some of the latest researches in classification and clustering tweets. Most of tweet's classification researches used in sentiment analysis. All tweets clustering researches used in topic and trends detection.

For future work, we would like to expand the scope of our comparisons and run the classifiers on tweets dataset considering Arabic language to have more representative inputs and thus better generalizable results. We focused on exploring hybrid approaches, where multiple models and techniques are combined to enhance the classification and clustering tweets.

8. REFERENCES

- [1] Manzoor, U., Baig, S. A., Hashim, M., & Sami, A., "Impact of social media marketing on consumer's purchase intentions: the mediating role of customer trust", 2, 2020., International Journal of Entrepreneurial Research, Vol. 3, pp. 41-48.
- [2] Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. 1, "The future of social media in marketing", 2020, Journal of the Academy of Marketing Science, Vol. 48, pp. 79-95.
- [3] Hernández-Fuentes, A., & Monnier, A., "Twitter as a Source of Information? Practices of Journalists Working for the French National Press", 5, 2020, Journalism Practice, Vol. 16, pp. 920-937.

- [4] Erskine, N., & Hendricks, S., "The use of Twitter by medical journals: systematic review of the literature", 7, 2021, Journal of medical Internet research, Vol. 23, p. e26378.
- [5] Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. 1, "Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods", 2022, Cognitive computation, Vol. 14, pp. 372-387.
- [6] Khan, A. Alsaedi and M. Z., "A study on sentiment analysis techniques of Twitter data", 2, 2019, International Journal of Advanced Computer Science and Applications, Vol. 10, pp. 361–374.
- [7] Mujahid, M., Lee, E., Rustam, F. and Washington, P.B., Ullah, S., Reshi, A.A., Ashraf, I., "Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19", 18, 2021, Applied Sciences, Vol. 11, p. 8438.
- [8] R. Jain, S. Bawa, and S. Sharma., "Sentiment Analysis of COVID-19 Tweets by Machine Learning and Deep Learning Classifiers", Singapore: Springer, Singapore, 2022, Advances in Data and Information sciences, pp. 329–339.
- [9] Chandrasekaran G, Hemanth J., "Deep learning and TextBlob based sentiment analysis for coron-avirus (COVID-19) using twitter data ", 1, 2022, International Journal on Artificial Intelligence Tools, Vol. 31, p. 2250011.
- [10] Singh, C., et al., "A Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews", 8, 2022, Applied Sciences, Vol. 12, p. 3709.
- [11] A. P., & Chiplunkar, N. N. , "A new big data approach for topic classification and sentiment analysis of Twitter data. Rodrigues", 2, 2022, Evolutionary Intelligence, Vol. 15, pp. 877-887.
- [12] A. P., & Chiplunkar, N. N., "Twitter sentiment analysis using machine learning for product evaluation", Coimbatore, India : s.n., 2020. Proc. 5th Int. Conf. on Inventive Computation Technologies. pp. 181–.
- [13] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel., "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification", 413, 2020, Information, Vol. 1.
- [14] Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P., "The impact of features extraction on the sentiment analysis", 2019, Procedia Computer Science, Vol. 152, pp. 341-348.
- [15] Alharbi, A. S. M., & de Doncker, E., "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information", 2019, Cognitive Systems Research, Vol. 54, pp. 50-61.
- [16] Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B., "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning", 6, 2019, Entropy, Vol. 21, p. 589.
- [17] Najafi, A., Gholipour-Shilabin, A., Dehkharghani, R., Mohammadpur-Fard, A., & Asgari-Chenaghlu, M., "ComStreamClust: a Communicative Multi-Agent Approach to Text Clustering in Streaming Data", 2022, Annals of Data Science, pp. 1-23.
- [18] Alomari, E., Mehmood, R., & Katib, I., "Road Traffic Event Detection Using Twitter Data, Machine Learning, and Apache Spark", s.l. : IEEE, 2019, In 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) , pp. 1888-1895.
- [19] Asgari-Chenaghlu, M., Nikzad-Khasmakhi, N., & Minaee, S., "Covid-Transformer: Detecting Trending Topics on Twitter Using Universal Sentence Encoder", 2020, <https://arxiv.org/pdf/2009.03947.pdf>.
- [20] Hernandez-Mendoza, M., Aguilera, A., Dongo, I., Cornejo-Lupa, J., & Cardinale, Y., "Credibility Analysis on Twitter Considering Topic Detection", 18, 2022, Applied Sciences, Vol. 12, p. 9081.
- [21] Khan, H. U., Nasir, S., Nasim, K., Shabbir, D., & Mahmood, A. , "Twitter trends: a ranking algorithm analysis on real time data", 2021, 113990, s.l. : Expert Systems with Applications, Vol. 164.
- [22] Octaria, O., Manongga, D., Iriani, A., Purnomo, H. D., & Setyawan, I., "Mining Opinion Based on Tweets about Student Exchange with Tweepy and TextBlob", 2022, In 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), pp. 102-106.
- [23] Chaudhri, A. A., Saranya, S. S., & Dubey, S., "Implementation paper on analyzing COVID-19 vaccines on twitter dataset using tweepy and text blob", 2021, Annals of the Romanian Society for Cell Biology, pp. 8393-8396.
- [24] Chandralekha, E., Rama, P., Benjamin, M. I. S. W., & Kumar, K. A., "Sentiment Analysis of Corona Vaccination on Twitter Data using Machine Learning Techniques", 2022, In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, pp. 708-712.
- [25] Manguri, K. H., Ramadhan, R. N., & Amin, P. R. M., "Twitter sentiment analysis on worldwide COVID-19 outbreaks", 2020, Kurdistan Journal of Applied Research, pp. 54-65.
- [26] Reynard, D., & Shirgaokar, M., "Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? Transportation research part D: Transport and environment", 2019, Vol. 77, pp. 449-463.
- [27] Lad, S., Mane, G., Padwal, A., & Dixit, M. 1, "Machine learning based sentiment analysis of Twitter data", 2018, In AIP Conference Proceedings, AIP Publishing LLC, Vol. 2494, p. 050007.
- [28] Zheng, X., & Sun, A., "Collecting event-related tweets from twitter stream", 2, 2019, journal of the Association for Information Science and Technology, Vol. 70, pp. 176-186.
- [29] Boon-Itt, S., & Skunkan, Y., "Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study", 4, 2020, jMIR Public Health and Surveillance, Vol. 6, p. e21978.
- [30] de Arriba Serra, A., Oriol Hilari, M., & Franch Gutiérrez, J., "Applying sentiment analysis on Spanish tweets using BETO", sep. 2021, international Conference of the Spanish Society for Natural Language Processing: Málaga, Spain : CEUR-WS. Org., In Proceedings of the Iberian Languages

Evaluation Forum (IberLEF 2021): co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2021), pp. 1-8.

- [31] Kharde, V., & Sonawane, P., “Sentiment analysis of twitter data: a survey of techniques”, 11, april 2016, International Journal of Computer Applications , Vol. 139, pp. 975 – 8887.
- [32] Resyanto, F., Sibaroni, Y., & Romadhony, A., “Choosing the most optimum text preprocessing method for sentiment analysis: Case: iPhone Tweets”, s.l. : IEEE, 2019. 2019 Fourth International Conference on Informatics and Computing (ICIC). pp. 1-5.
- [33] Saputra, D. D., Gata, W., Wardhani, N. K., Parthama, K. S., Setiawan, H., Budilaksono, S., ... & Novianti, D., “Optimization Sentiments of Analysis from Tweets in myXLCare using Naïve Bayes Algorithm and Synthetic Minority Over Sampling Technique Method”, 1 Feb. ,2020, s.l. : IOP Publishing In Journal of Physics: Conference Series. Vol. 1471 no. 1, p. 012014.
- [34] Matt K., Yu S., Nicholas K., and Kilian W., “From word embeddings to document distances”, 2015, s.l. : in International conference on machine learning, pp. 957–966.
- [35] Algorithms for Clustering Data; Prentice-Hall. Jain, A.K. and Dubes, R.C. s.l. : Upper Saddle River, NJ, USA, 1988.
- [36] Modeling Topics Technical Report; Carnegie Melon University: Pittsburgh,. Gimpel, K. s.l. : PA, USA, 2006.
- [37] An introduction to latent semantic analysis. Landauer, T.K., Foltz, P.W. and Laham, D. s.l. : Discourse Process. 1998, 25, 259–284.
- [38] Nonnegative Matrix Factorization: A Comprehensive Review. Wang, Y.X. and Zhang, Y.J. s.l. : IEEE Trans. Knowl. Data Eng. , 2012, Vols. 25, 1336–1353.