



Performance Evaluation of Sentiment Analysis Using Machine Learning Models

Alexander Ntino, Souley Boukari, Sani Abba, Monday Simon
Department of Mathematical Sciences,
Abubakar Tafawa Balewa University, Bauchi, Nigeria.

ABSTRACT

Sentiment analysis is the process of using natural language processing and machine learning techniques to determine the emotional tone or sentiment expressed in a piece of text, such as positive, negative, or neutral. Many techniques were used to evaluate similar models but the developed model utilizes natural language preprocessing, text mining and sentiment analysis techniques to analyze Twitter datasets using deep learning algorithms. Pre-trained word embedding's, such as Word2Vec were utilized to capture semantic relationships, while TF-IDF vectorization was employed to convert text into numerical features. The dataset was split into training and testing sets, and machine learning models, LSTM and Naive Bayes are trained and evaluated based on accuracy metrics. The performance of LSTM, and Naive Bayes were compared. LSTM has a performance of 0.937, and Naive Bayes 0.795. The results indicate that, with an accuracy of 0.937 in predicting sentiment analysis on Twitter, the LSTM model outperforms Naive Bayes that was developed and tested using the same dataset and system configuration. The developed Model demonstrates superior performance compared to the baseline model in terms of accuracy, precision and Recall with the performance improvement of 0.142 in accuracy. We recommend this model to be deployed on social media for sentiment analysis.

ARTICLE INFO

Article History

Received: August, 2023

Received in revised form: August, 2023

Accepted: September, 2023

Published online: September, 2023

KEYWORDS

Machine Learning, Long Short Term Memory, Naive Bayes.

INTRODUCTION

With the rapid development of information and communication technologies, Rich user-generated content (UGC) has proliferated on social networking sites and blogs, which have become major sources of UGC. Information system academics have noted that as a result of the growth of online search engines, e-commerce, and social media, business intelligence has evolved from the analysis of historically structured business transaction data to the study of unstructured and real-time data. The Twitter microblog is one of the most well-liked platforms currently available for social media, having been embraced by both customers and businesses (Srinivas et al., 2021). These days' social media, blogs, and networks produce huge information. People share their thoughts on

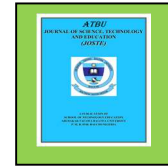
various events and issues. This bulk information is used for opinion and decision-making in a business. Twitter generates massive short tweets and because of that, the tweet size is reduced to 140 characters. Opinion and Sentiment Mining are exceptionally helpful to get the opinion of the users or to analyze their sentiments (Srinivas et al., 2021). Over half a billion Twitter messages(tweets) per day are being recorded, and a recent customer engagement technology study reported that Twitter is among the top three most used social media platforms by US hotel and restaurant operators (Philander & Zhong, 2016). Online Complaint Management System provides a time-efficient and corruption-free way of resolving the problems faced by the public. The system makes complaints easier to coordinate, monitor, track, and resolve, and provides the

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



company with an effective tool to identify and target problem areas, monitor complaints handling performance, and make business improvements. Several complaint management systems facilitate the storage and retrieval of complaint data and very few consider advancing its processes with recent Artificial Intelligence technology and in real-time. Artificial Intelligence technics are helpful in the complaint resolution process, understanding the attitudes of involved parties and reasoning about them, in particular, using Belief–desire–intention mode, and very integral to this is the dictation of spam. Several real-time techniques can be used for sentiment analysis, which involves determining the emotional tone behind a piece of text, such as a tweet or a review (Rodrigues et al., 2022).

Real-time refers to a mode of computer operation in which events or data processing occur immediately as they happen, without any noticeable delay. In other words, a system can process and respond to inputs or events as they occur in the actual moment, they take place. Real-time systems are designed to provide timely and instantaneous responses, often within milliseconds or microseconds, depending on the context. With Artificial Intelligence (AI) in real-time, complaint management will be easier, faster, more accurate, more instantaneous, more convenient, and can provide a level of responsiveness than human agents. AI can speed up a complaint management process by figuring out people's intent and routing them to the right agent automatically. The technology learns words and phrases commonly used in inquiry and then predicts the nature, tone, and other details of the complaint (Sentiment). Social media is an essential piece of any business marketing strategy. Considered as a sub-field of Artificial Intelligence (AI), machine learning enables computer systems to automatically learn from data patterns rather than being programmed following particular rules (Rodrigues et al., 2022).

Social platforms help connect customers, increase awareness about the brand, and boost leads and sales. About 52 percent of social media marketers believe social media positively influences their company's revenue and sales (El Barachi, & Oroumchian, 2021) because

more than 3 billion people now use social networks worldwide, these people interact with brands on social media. More people follow brands on social media than celebrities, according to Sherpa Marketing. 80 percent of users follow at least one brand on Instagram alone. If social media is not being used, brands will be losing out on a quick, affordable, and efficient way to connect with roughly half of the world's population. To learn about user attitudes or feelings in various situations when pandemics are occurring around the world, some academics have been working on sentiment analysis on various social media data, particularly on Twitter. (Kausar, & Nasar, 2021)

In general, opinion mining has been used to know about what people think and feel about their products and services on social media platforms. Every day, millions of individuals express their ideas on various parts of life. Spurred by that growth, companies and media organizations are increasingly seeking ways to mine information. It requires efficient techniques to collect a large amount of social media data and extract meaningful information from them (Trupthi, Pabboju, & Narasimha, 2017).

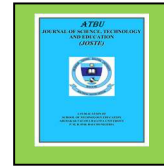
Sentiment analysis is a textual mining technique used to extract subjective information from social media platforms like Twitter to understand social sentiments about brands and products (Srinivas et al., 2021). Huge unstructured data is available in many forms like tweets, reviews or news articles etc. which can be classified as positive, neutral or negative polarity according to the sentiment that is expressed by them (Trupthi, & Narasimha, 2017). According to the work of Irmayani, (2021) a system was developed which used a classifier from a large data set at the start up that can perform classification of tweets based on the classifier built. Performing sentimental analysis on the tweets based on naïve Bayes' classifier trained from a large data set and provide time variant analytics based on the results obtained. Training should be performed on a large data set which is the major criteria to get efficient classifier. Other section of the paper is as follows: Section 2. Problem Statement. Section, 3. Related works, Section 4. Methodology, Section 5.Result,Section

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



6. Conclusion and Section 7. Recommendation and Future Research.

PROBLEM STATEMENT

The approach by (Irmayani, et al., 2021) performs real-time sentimental analysis on the tweets that are extracted from Twitter and provides time-based analytics to the user using Naïve Bayes classification approach. The key features of the system are the training module which is done with the help of Hadoop and MapReduce, Naive Bayes, Time Variant Analytics, and the Continuous Learning System-based Classification. Despite the advancements in sentiment analysis, there is a lack of comprehensive studies that evaluate the performance of different machine learning models in a real-time context, there is a critical gap in the ability to perform real-time sentiment analysis effectively and efficiently. Existing sentiment analysis models, particularly those relying on traditional machine learning approaches, often struggle to provide timely insights due to their inability to handle the rapidly changing landscape of textual data. These models typically lack the capacity to process streaming data sources with low latency, a crucial requirement for various applications.

The existing system has some limitations. When classifying the sentence, words are taken individually rather than the sentence in total. The semantic meaning is neglected that might be present between the words. The second Limitation is the system may not provide the actual intended meaning to the user. There might be some sort of sarcasm present in the sentences which the system ignores. The challenge lies in developing a robust and efficient system that can accurately analyze the sentiment behind this dynamic and unstructured textual data in real time. Additionally, the Naïve Bayes algorithm as used in the existing studies makes the assumption that all features are independent of one another, which is frequently false in practical application. If the features are correlated, this may result in inaccurate classification results (Arar & Ayan, 2017). Therefore, as suggested in (Irmayani et al., 2021), further research is expected to use different

algorithms to obtain more accurate results in the analysis of opinions.

Traditional sentiment analysis approaches have made significant strides in automating sentiment classification, but they often fall short when applied to real-time scenarios. LSTM, a type of recurrent neural network (RNN), has shown promise in capturing long-range dependencies in sequential data and is particularly well-suited for processing text (Jazib, Tariq, & Mahmood, 2022).

This paper leverages on the work of (Irmayani et al., 2021), to develop an improved model and evaluate the performance of the model with the benchmark paper in terms of accuracy.

RELATED WORKS

Mining textual data to extract sentiment has been a challenging aspect of Natural Language Processing (NLP) due to the unstructured form of text and the discovery of information patterns in the text; most machine learning algorithms have been applied to numerical and categorical variable on structured data frames. Thus, the full dependence on NLP and complex techniques and algorithms to manipulate textual data, discover useful pattern of information. These techniques deviate marginally from legacy languages perceived by computers and machines. Many researches have been conducted and practically applied to real world problems in the past 2 decades. The basic application of sentiment analysis lies in gathering the opinion of people. Such opinions are precursors to many business decisions. Similar to the stop-words, sentiment analysis domain depends on list of words that describe the effect of the writer, Sentiment, Text Analytics, Anonymous, Sentiment-Analyzer, Text-Processing, Machine Learning Analyzer, etc. explains how ANEW list is used to classify the opinions of users as negative, neutral or positive applied the sentiment analysis concepts to topics rather than the actual opinions and gleaned how the discussion would follow the sentiment of a topic (Chouhan et al., 2023).

(Gandhi, & Karthick, 2021) stated that, the initial stage in movie review is importing the IMDB movie review dataset, which explains how

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



the used Deep Learning models are implemented for CNN and LSTM approaches. The review data was then extracted using natural language processing tools like word2vec and stop words. A convolutional layer with a 1- dimensional matrix was added before global max pooling was used to produce the final product. The outcome for the CNN test classification will result in the movie review after conducting a fully connected layer.

Kanojia & Joshi, (2023) examined the views of democratic voters and their feelings over the presidential elections examined the long-term application of sentiment analysis in politics so that the electorate may gain a long-term perspective on the behavior of the elected officials. Other projects have as a long-term goal the clarification of politicians' positions. Shanmugavadivel, & Subramanian, (2022) suggested a method that uses machine learning to analyze sentiment and identify offensive words in low resource code mixed Tamil and English data. (Cheong & Lee, 2011) tackled the problem of finding content related to terrorism. The study examined ways to gather such information from twitter data by harvesting user sentiments from the public. More work has been done in the field of twitter harvesting by (Rosenthal, & Nakov, 2019). Twitter data was also used to build a corpus that could be specific to twitter. Several complain management systems facilitate the storage and retrieval of complain data and very few consider advancing it processes with recent Artificial Intelligence technology. The two broad classes of sentiment analysis approaches are those applying sentiment lexicons and scoring algorithms, and those using machine-learned models for classification (Zimbra, & Chen, 2018). Furthermore, two classes of machine-learned models are considered, systems pre- trained on a general sentiment analysis corpus requiring no additional calibration for application, and others trained within each domain of application. General-purpose sentiment analysis systems are convenient and ready to use but lack exposure to domain-specific expressions of sentiment, which may limit their performance (Zimbra et al., 2018).

With Artificial Intelligence (AI), sentiment analysis will be easier, faster, more accurate and more convenient and can provide a level of

responsiveness than human agents. AI can speed up a complaint management process by figuring out people's intent and routing them to the right agent automatically (Carneiro, & Neves, 2014). The technology learns words and phrases commonly used in inquire and then predicts the nature, tone and other details of the complaint or sentiment. (Rodrigues et al., 2022) developed a system that can determine whether a tweet is "spam" or "ham" and evaluate the emotion of the tweet. The extracted features after pre-processing the tweets are classified using various classifiers, namely, decision tree, logistic regression, multinomial Naive Bayes, support vector machine, random forest, and Bernoulli Naive Bayes for spam detection

Trupthi, & Narasimha, (2017) proposed sentiment analysis on twitter using streaming API and evaluate it using Naïve Bayes. The fact that the analysis is done real time is the major highlight of this paper. However, the major limitation of this study is that, when classifying the sentence, words are taken individually rather the sentence in total. The semantic meaning is neglected that might be present between the words.

Similarly, (Kamyab, et al., 2018) proposed Sentiment Analysis on Twitter using text Mining Approach to the Afghanistan Status Reviews. The proposed approach was evaluated using Machine Learning Classifiers. The result of sentiment analysis indicated that the number of tweets classified, 34% of them were classified as negative, 35.1% as neutral, and 30.9% as positive. A limitation of comments. Thus, the results need to be further tested with a large amount of comment written in other languages.

Additionally, Sasikala & Sheela, (2018) proposed a sentiment Analysis of Online Food Reviews using Customer Ratings using Naïve Bayes and Regression. The proposed system has produced a very good result using the score ratings. The limitation of this system is, it works better only for the open sentiments like rating or scores. The results were not promising for hidden sentiments.

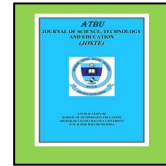
More so, Ashi, et al., (2018) proposed pre-trained Word Embedding's for Arabic Aspect-Based Sentiment Analysis of Airline Tweets using SVM. The Results show that such word

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



embedding features yet simple are powerful enough to achieve state of the art performance in performing ABSA for the Arabic language however, the research fail to perform multi-label classification ABSA using Twitter where each tweet can fall into one or more classes/labels.

Ogudo & Nestor, (2019) proposed sentiment Analysis Application and Natural Language Processing for Mobile Network Operators' Support on social media using Machine learning & hierarchical clustering. The orientation of the study in this paper creates a bi-directional link between customers or followers and the MNO with an accuracy of 96.516%

Rajput, (2020) proposed a comprehensive study on Natural Language Processing, Sentiment Analysis and Clinical Analytics lexicons and scoring algorithms. However, the research fails to identify how sentiment analysis and NLP techniques can be helpful to cyber bullying.

Based on the review, several approaches for sentiment analysis have been proposed in the literature, however, very few consider advancing its processes with recent Artificial Intelligence technology Sentiment analysis, also known as opinion mining, offers a wide range of real-time applications but has been plagued by numerous research restrictions (Hazarika, Konwar, Deb, & Bora, 2020). With Artificial Intelligence (AI), sentiment analysis will be easier, faster, more accurate and more convenient and can provide a level of responsiveness than human agents. AI can speed up a complaint management process by figuring out people's intent and routing them to the right agent automatically (Irmayani et al., 2021). The technology learns words and phrases commonly used in inquiry and then predicts the nature, tone and other details of the complaint (Sentiment). Hence, we focus our review to those using

machine-learned models for classification algorithms due to the promising results achieved by these techniques.

The existing system developed by (Trupthi, Pabboju, & Narasimha, 2017) perform real time sentimental analysis on the tweets that are extracted from Twitter and provides time-based analytics to the user using Naïve Bayes classification approach. The key features of the system are the training module which is done with the help Hadoop and MapReduce, Classification based on Naïve Bayes, Time Variant Analytics, and the Continuous Learning System. The fact that the analysis is done real time is the major highlight of their research.

METHODOLOGY

In this study, the forward chaining inference mechanism methodology is used, which stipulates the step-by-step procedure undertaken to arrive at the classification of positive, negative and neutral comments base on the retrieved sentiment datasets from twitter. These sections include the following: analysis of the present (conventional) system; identification of problems of the current system; analysis of the proposed system; and system implementation tools of present system and its feasibility.

Existing Model

The Model proposed by Irmayani et al., (2021) extracts the data from Social Network Site services which is done using Streaming API of twitter. The extracted tweets are loaded into Hadoop and it is been preprocessed using map reduce. This task is followed by classification which uses NLP and machine learning techniques. The classification used here is uni-word Naïve Bayes' classification. The architecture of the existing methodology is presented in Figure 1.

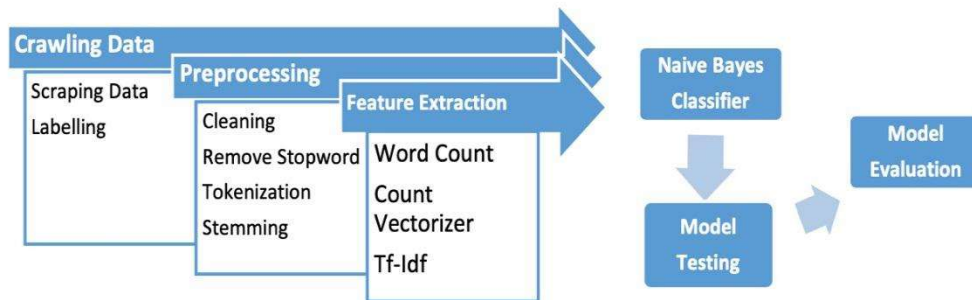


Figure 1. Diagram of Research Method (Irmayani et al., 2021).

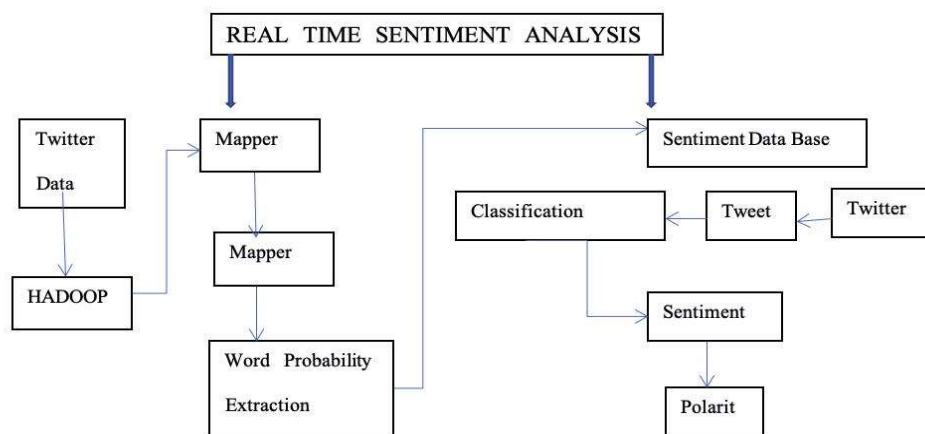


Figure 2. Bloch Diagram of the Existing system by (Trupthi, Pabboju, & Narasimha, 2017) .

Several existing systems store old tweets and perform sentiment analysis on them, giving results on old data and using up a lot of space but in this system, the tweets do not need to be stored which is cost effective as no storage space is needed. Also, all the analysis is done on tweets in real time, So, the user is assured that they, getting new and relevant results. However, the existing system has some limitations. The first limitation is being Uni-gram Naïve i.e. training of the data was done based on word probabilities and used the same for classification. Future enhancement to this work might be to use n-gram classification rather than limiting to uni-gram which will require pattern filtering on Hadoop. When classifying the sentence, words are taken individually rather than the sentence in total. The semantic meaning is neglected that might be present between the words. It might be possible to

build a system that can perform sentiment analysis in all languages. The third limitation is that the system may not always provide the actual intended meaning to the user. There might be some sort of sarcasm present in the sentences which the system ignores.

As seen in the literature many attributes make NB user-friendly, like its simplicity, low sample requirement, and ease of implementation Naive Bayes assumes that all features are independent, which may not be true in real-world scenarios. This can lead to inaccurate predictions (Derbentsev, & Hostryk, 2023). Additionally, the Naive Bayes relies on the "bag of words" assumption, which means it ignores the order and context of words in a text. This can result in the loss of important information for sentiment analysis (Derbentsev et al., 2023). Naive Bayes can have difficulty classifying text that contains

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved

rare words or phrases. This is because the model calculates the probability of each word or phrase in the test data based on how often it appeared in the training data. If a word or phrase does not appear frequently in the training data, the model may assign it a very low probability, even if it is actually important for classification (Ferdiansyah, 2022). Therefore, as suggested in the base paper (Irmayani et al., 2021), further research is expected to use different algorithms to obtain more accurate results in the analysis of opinions. Hence, in this study, we proposed to implement a

classification algorithms base on deep learning architecture to further address the aforementioned problem witness in Naïve Bayes' algorithm.

Proposed Method

The proposed approach receives data from twitter and calculate sentiment scores using machine learning algorithm, for each tweet and categorized them as positive or negative. The abstract view of the proposed approach is shown in Figure 3. Consisting of four main modules as described below:

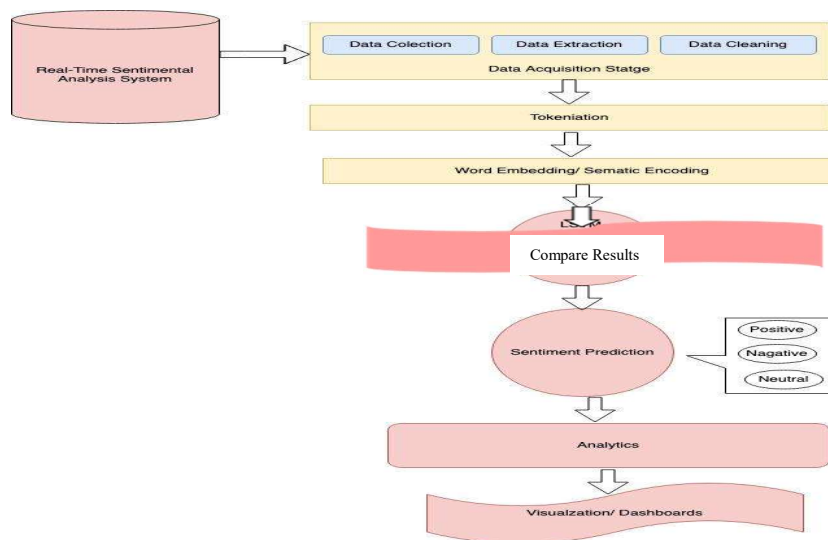


Figure 3: Architecture of the proposed sentiment analyzer

The proposed system extracts the data from SNS services using the Streaming API of Twitter, Public Firehose Stream, which provides access to the full stream of all public tweets, including a much higher volume of data, The extracted tweets are loaded into Hadoop as part of the real-time sentiment analysis and analytics system primarily for batch processing and distributed storage and it is being preprocessed using map-reduce and also Machine Learning Model Training. This task is followed by classification which uses NLP and machine learning techniques. Consider the number of all positive tweets, positive words, and negative

words from our training phase. Then the model will calculate the probability of a tweet being positive.

Figure. 3 lists an approach we used for integrating natural language preprocessing, text mining, and sentiment analysis techniques to analyze the Twitter database on deep learning algorithms. Below is a description of the steps:

Data Acquisition

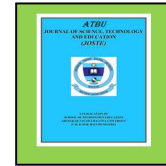
The dataset is based on data from the following two sources: Niek Sanders' University of Michigan Sentiment Analysis contest on Kaggle Twitter Sentiment Corpus. The Twitter Sentiment Analysis Dataset contains 60400 classified tweets, each row is marked as 1 for positive

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



sentiment, 0 for neutral sentiment and -1 for negative

Data preprocessing and cleansing

Natural language preprocessing used to clean the tweets such as removing stop words to allow us execute all of the cleaning tasks, removing all URL's, punctuations, symbols, numbers, stop words and non-English tweets. This preprocessing step will reduce the volume of our dataset. After the preprocessing part, all tweets were be split into tokens, and those words were looked up in each sentence and scores will be calculated.

Feature Extraction and Tokenization

Feature extraction/reduction is usually used when the original data was very different. Feature engineering is the careful preprocessing into more meaningful features, even if you could have used the old data. It is frequently regarded as applied machine learning and is effective for enhancing the performance of machine learning algorithms. Algorithms for machine learning and data analytics can be more efficiently computed by focusing on the key features and condensing the feature set. All classifier work with feature set, which is a simple dictionary mapping feature name to feature value. In order to do so, the model will be used to generate a feature vector of each tweet, where the occurrence of each word is used as a feature and Term Frequency Inverse Document Frequency is used as feature value. To generate feature vectors from tweet several basic Information Retrieval (IR) techniques are applied.

Tokenization: is the process of splitting the input text into a sequence of individual units called tokens. Tokenization can be applied at different level such as paragraph, sentence, and word level. Since we need words as feature vector, therefore, word level tokenization is performed using non- letter characters.

Word Embedding

Word embedding is a fundamental technique in natural language processing (NLP) and deep learning that represents words as numerical vectors in a continuous vector space. It

transforms words from discrete symbols into dense, real-valued vectors. Word embeddings capture semantic relationships between words and allow machines to understand and work with words in a way that is more mathematically meaningful.

Word2Vec is employed in the neural networks to learn word vectors from a large text corpus. Word2vec is a popular model that represents words in vector form, capturing their semantic meaning and enabling tasks such as text classification, analogy finding, and syntactic and semantic analysis of words (Sivakumar, & Haritha, 2020).

Sentiment Prediction

After the developed model is trained, the score of each sentence in the source tweets is calculated with the total weight of each term in the relevant sentences. The sentiment is classified into positive, neutral and negative based on the prediction. The sentiment is classified based on the keywords of the intended topic to be analyzed.

Experimental Setup and System Implementation

The experiment was conducted on Python Jupyter Notebook 6.5.4 running a Microsoft Window 10 with 64-bit Operating System, 8GB (RAM) and Intel(R) Core (TM) i7-4000M @ 2.4 GHz. To implement the proposed model, this research uses a Twitter Sentiments textual dataset, the dataset has three sentiments namely, negative, neutral, and positive, it contains two fields for the tweet and label. The dataset is based on data from the following two sources: Niek Sanders' University of Michigan Sentiment Analysis contest on Kaggle Twitter Sentiment Corpus.

Additionally, all these Tweets and Comments were extracted using their Respective Apis Tweepy and PRAW. All the Tweets and Comments are Cleaned using Pythons and also, NLP with a Sentimental Label to each ranging from -1 to 1. Twitter_Data.csv Dataset has around 60,400 Tweets along with Sentiment Labels, Each Dataset has two columns, the first column has the cleaned tweets and Comments and the Second one indicates its Sentimental Label.

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved

Table 1: Category and Classes of the Twitter Datasets

Sentiment	Class
-1	Negative Neutral Positive
0	
1	

RESULT

The prediction performance for the suggested models on the three distinct sentiment classes is shown in this section following dataset experimentation. Python Jupyter Notebook 6.5.4 was used to develop all algorithms, and CountVectorizer from scikit-learn was used for data preprocessing to turn the preprocessed tweet texts into a bag-of-words representation.

LSTM Result

The LSTM classification model demonstrates strong performance in predicting sentiment classes, with high Area under Curve (AUC) scores for positive, neutral, and negative sentiments. The graphical representation in Figure 3 graphically shows the performance of the LSTM classification model at all classification thresholds.

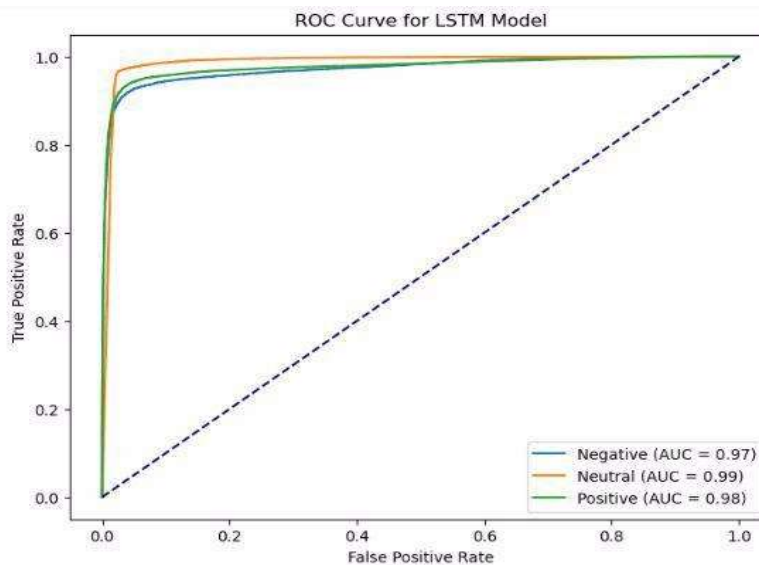


Figure 4: ROC Curve Graph for LSTM Model

The performance of the LSTM classification model in accurately predicting each sentiment class at all classification thresholds is

graphically displayed in Figure 4, with AUCs for positive, neutral, and positive sentiment of 97%, 99%, and 98%, respectively.

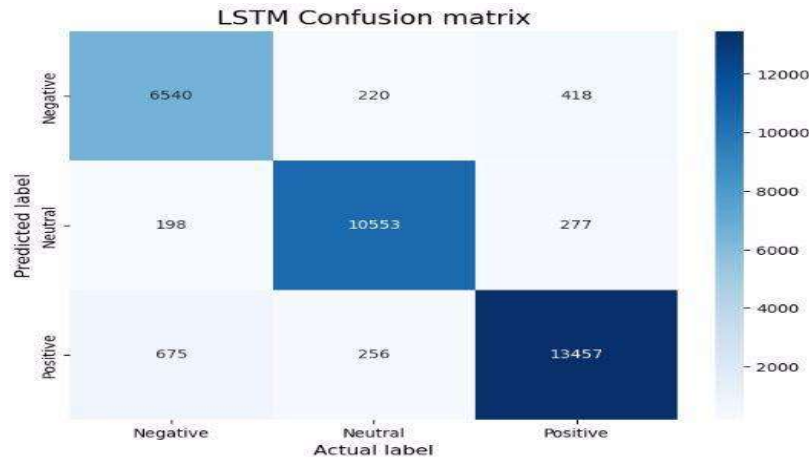


Figure 5: The result output of confusion matrix for LSTM model

The effectiveness of the LSTM classification model in predicting based on the predicted label relative to the actual label is shown in Figure 5. The confusion matrix above demonstrates that, for each class, the model was able to predict true positives with a higher value than false positives, indicating that it had trained more effectively.

Naïve Bayes Result

The Naïve Bayes classification model demonstrates strong performance in predicting sentiment classes, with high AUC scores for positive, neutral, and negative sentiments. The graphical representation in Figure 6 graphically shows the performance of the Naïve Bayes classification model at all classification thresholds.

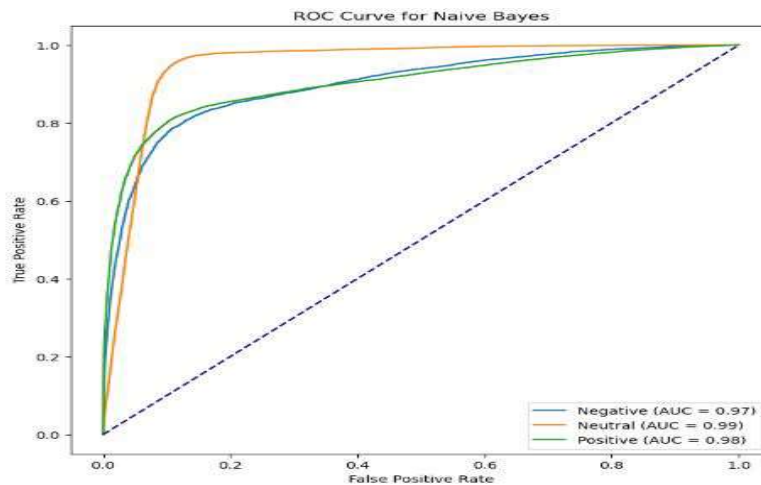


Figure 6: ROC Curve Graph for Naïve Bayes Model

The accuracy with which the Naïve Bayes classification model predicts each sentiment class at all classification thresholds is

graphically displayed in Figure 6. Correspondingly with AUCs of 97%, 99%, and 98% for positive, neutral, and positive sentiment.

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved

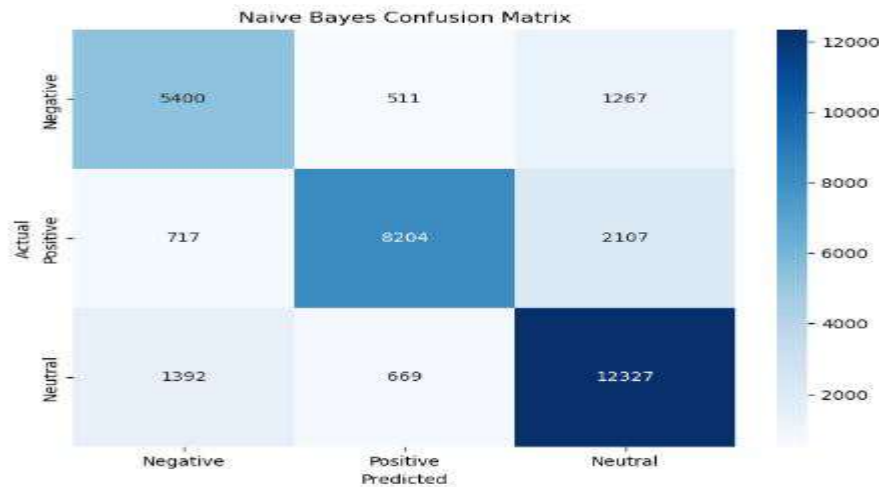


Figure 7: The result output of confusion matrix for Naïve Bayes model

The performance of the Naïve Bayes classification model in predicting based on the predicted label compared to the actual label is depicted in Figure 7. As can be seen from the confusion matrix above, for each class the model was able to predict true positives with a higher value than false positives, indicating that it had trained more effectively.

Comparison Result

To determine the best model for Twitter sentiment analysis, the three developed models were tested using test set data after training, and their accuracy was compared. As been represented below:

Table 2: The result comparison table

	lstm_model	nb_model
evaluation		
Accuracy	0.937289	0.795576
Precision	0.940490	0.800742
Recall	0.934835	0.795576

Table 3: The result Comparison using Accuracy

Accuracy	
evaluation	
lstm_model	0.937289
nb_model	0.795576

The comparison results for the three developed models based on accuracy are displayed in table 2 and 3 above. Nonetheless, compared to other models, the LSTM model was able to predict Twitter sentiment more accurately, at 93.7%.

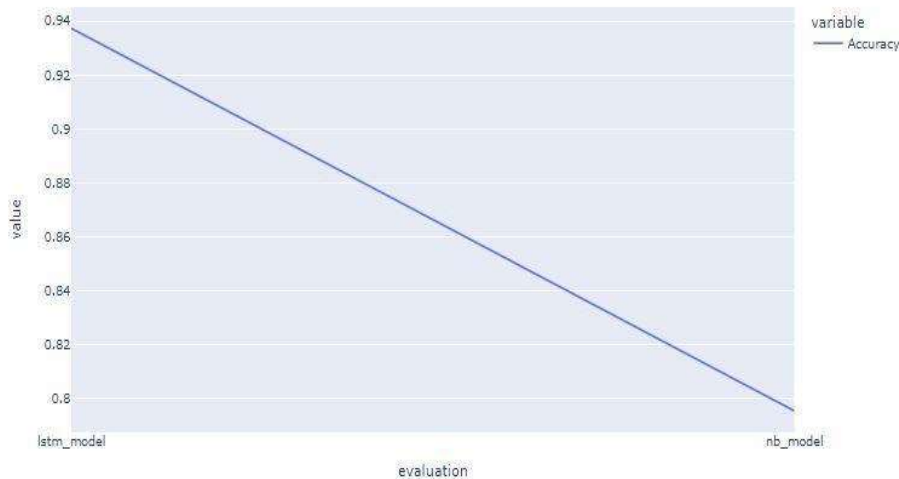


Figure 8: graphical representation of comparison result

The graphical representation of the comparison result in Figure 8 indicates that, with an accuracy of 0.937, the LSTM model outperformed the Naïve Bayes model.

DISCUSSION OF RESULT

The results indicate that, with an accuracy of 0.937 in predicting sentiment analysis on Twitter, the LSTM model outperforms the Naïve Bayes model that were developed and tested using the same dataset and system configuration. However, when compared to the baseline model, the suggested model outperforms the baseline model in terms of accuracy. Thus, the LSTM model demonstrated superior performance in sentiment analysis on Twitter compared to alternative models and a baseline model. The high accuracy achieved by the LSTM model indicates its effectiveness in capturing and understanding sentiment patterns in the Twitter dataset used in the study. The statement suggests that the research goals were met successfully, emphasizing the accomplishment of the project objectives.

CONCLUSION

This study delves into the realm of sentiment analysis on Twitter, leveraging advanced techniques such as Naïve Bayes, and Long Short-Term Memory (LSTM) networks, the

study scrutinizes the nuances of sentiment classification. The comparison is not merely confined to accuracy metrics but extends to considerations of model robustness and performance across different sentiment categories. The results indicate that, with an accuracy of 0.925 in predicting sentiment analysis on Twitter, the LSTM model outperforms the other two models that were developed and tested using the same dataset and system configuration. However, when compared to the baseline model, the suggested model outperforms the baseline model in terms of accuracy. Thus, the LSTM model demonstrated superior performance in sentiment analysis on Twitter compared to alternative models and a baseline model. The high accuracy achieved by the LSTM model indicates its effectiveness in capturing and understanding sentiment patterns in the Twitter dataset used in the study. The statement suggests that the research goals were met successfully, emphasizing the accomplishment of the project objectives. In conclusion, the study explored into sentiment analysis on Twitter using various machine learning models has provided valuable insights into the effectiveness of these models in predicting sentiment classes, demonstrating distinct strengths and limitations, contributing to a nuanced understanding of their applicability in the context of sentiment analysis.

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



The major limitations of this research are that, the research fails to fully capture the diversity of sentiments expressed in the broader population due to potential sampling biases inherent in social media platforms and does not explicitly account for temporal dynamics, the models may not adequately capture shifts in sentiment patterns over time. Lastly, other models or hybrid approaches that were not considered may offer alternative perspectives and potentially better performance.

RECOMMENDATION AND FUTURE RESEARCH

The study recommends, that investigating hybrid approaches that combine the strengths of different models could be a fruitful avenue. Ensemble methods or hybrid architectures may leverage the unique features of each model, potentially leading to improved overall performance. Also, Future research should explore methods for capturing temporal dynamics, such as sentiment trends over time and the impact of specific events on sentiment patterns. Future research could benefit from incorporating time-series analysis or dynamic modeling techniques to capture the influence of evolving events and trends on sentiment patterns over time. Future research could investigate the efficacy of exploring additional models or hybrid approaches could offer alternative perspectives and potentially lead to improved performance for analyzing sentiment in social media data.

REFERENCES

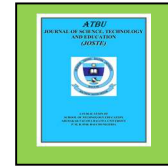
- Arar, Ö. F., & Ayan, K. (2017). A feature dependent Naive Bayes approach and its application to the software defect prediction problem. *Applied Soft Computing*, 59, 197-209.
- Ashi, M. M., Siddiqui, M. A., & Nadeem, F. (2018). *Pre-trained word embeddings for Arabic aspect-based sentiment analysis of airline tweets*. Paper presented at the International Conference on Advanced Intelligent Systems and Informatics.
- Carneiro, D., Novais, P., Andrade, F., Zeleznikow, J., & Neves, J. (2014). Online dispute resolution: an artificial intelligence perspective. *Artificial Intelligence Review*, 41(2), 211- 240.
- Cheong, M., & Lee, V. C. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, 13(1), 45-59.
- Chouhan, K., Yadav, M., Rout, R. K., Sahoo, K. S., Jhanjhi, N., Masud, M., & Aljhdali, S. (2023). Sentiment Analysis with Tweets Behaviour in Twitter Streaming API. *Comput. Syst. Sci. Eng.*, 45(2), 1113-1128.
- Derbentsev, V. D., Bezkorovainyi, V. S., Matviychuk, A. V., Pomazun, O. M., Hrabariev, A. V., Hostryk, A. M. (2023). *A comparative study of deep learning models for sentiment analysis of social media texts*. Paper presented at the CEUR Workshop Proceedings.
- El Barachi, M., Alkhatib, M., Mathew, S., & Oroumchian, F. (2021). A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. *Journal of Cleaner Production*, 312, 127820.
- Ferdiansyah, I. F. (2022). *Comparative Study of Hoax Detection using Support Vector Machine Algorithm, Naive Bayes, Random Forest and K-Nearest Neighbor*. Paper presented at the 2022 13th International Conference on Information and Communication Technology Convergence (ICTC).
- Gandhi, U. D., Malarvizhi Kumar, P., Chandra Babu, G., & Karthick, G. (2021). Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). *Wireless Personal Communications*. doi:10.1007/s11277-021- 08580-3
- Godara, J., Aron, R., & Shabaz, M. (2022). Sentiment analysis and sarcasm

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



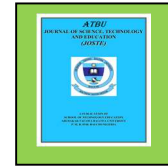
- detection from social network to train health-care professionals. *World Journal of Engineering*, 19(1), 124-133. doi:10.1108/WJE-02-2021-0108
- Hazarika, D., Konwar, G., Deb, S., & Bora, D. J. (2020). Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing. *ICRMAT*, 24, 63-67.
- Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., Zhang, J. (2021). *Lstm based sentiment analysis for cryptocurrency prediction*. Paper presented at the Database Systems for Advanced Applications: 26th International Conference, DASFAA 2021, Taipei, Taiwan, April 11–14, 2021, Proceedings, Part III 26.
- Irmayani, D., Edi, F., Harahap, J. M., Rangkuti, R. K., Ulya, B., & Watrionthos, R. (2021). *Naives Bayes algorithm for twitter sentiment analysis*. Paper presented at the Journal of Physics: Conference Series.
- Jazib, A., Tariq, W., & Mahmood, M. (2022). *Sentiment Analysis using Ensemble Classifier for Entrepreneurs based on Twitter Analytics*. Paper presented at the 2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST).
- Kamyab, M., Tao, R., Mohammadi, M. H., & Rasool, A. (2018). *Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews*. Paper presented at the Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality.
- Kanojia, D., & Joshi, A. (2023). Applications and Challenges of Sentiment Analysis in Real-life Scenarios. *arXiv preprint arXiv:2301.09912*.
- Kausar, M. A., Soosaimanickam, A., & Nasar, M. (2021). Public sentiment analysis on Twitter data during COVID-19 outbreak. *International Journal of Advanced Computer Science and Applications*, 12(2).
- Khrais, L. T. (2020). Role of artificial intelligence in shaping consumer demand in E-commerce. *Future Internet*, 12(12), 226.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications*, 82(3), 3713-3744.
- Lee, E., Rustam, F., Washington, P. B., El Barakaz, F., Aljedaani, W., & Ashraf, I. (2022). Racism detection by analyzing differential opinions through stacked analysis of tweets using stacked ensemble gcr-nn model. *IEEE Access*, 10, 9717-9728.
- Neri, F., Aliprandi, C., Capecci, F., Cuadros, M., & By, T. (2012). *Sentiment analysis on social media*. Paper presented at the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
- Ogudo, K. A., & Nestor, D. M. J. (2019). *Sentiment analysis application and natural language processing for mobile network operators' support on social media*. Paper presented at the 2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD).
- Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55(2016), 16-24.
- Radiuk, P., Pavlova, O., & Hrypynska, N. (2022). An ensemble machine learning approach for Twitter sentiment analysis.
- Rahmaniar, W., & Hernawan, A. (2021). Real-time human detection using deep learning on embedded platforms: A review. *Journal of Robotics and Control (JRC)*, 2(6), 462-468.
- Rajput, A. (2020). Natural language processing,

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved



- sentiment analysis, and clinical analytics. In *Innovation in Health Informatics* (pp. 79-97): Elsevier.
- Revathy, G., Alghamdi, S. A., Alahmari, S. M., Yonbawi, S. R., Kumar, A., & Haq, M. A. (2022). Sentiment analysis using machine learning: Progress in the machine intelligence for data science. *Sustainable Energy Technologies and Assessments*, 53, 102557.
- Rodrigues, A. P., Fernandes, R., A. A., B. A., Shetty, A., K. A., Shafi, R. M. (2022). Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques. *Computational Intelligence and Neuroscience*, 2022, 5211949. doi:10.1155/2022/5211949
- Rosenthal, S., Farra, N., & Nakov, P. (2019). SemEval-2017 task 4: Sentiment analysis in Twitter. *arXiv preprint arXiv:1912.00741*.
- Sasikala, P., & Sheela, L. M. I. (2018). Sentiment analysis of online food reviews using customer ratings. *International Journal of Pure and Applied Mathematics*, 119(15), 3509-3514.
- Services, D. W. (2019). What is Artificial Neural Network – Structure, Working, Applications. Retrieved from <https://data-flair.training/blogs/artificial-neural-network/>
- Shanmugavadivel, K., Sathishkumar, V., Raja, S., Lingaiah, T. B., Neelakandan, S., & Subramanian, M. (2022). Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. *Scientific Reports*, 12(1), 21557.
- Sisodia, D. S., Vishwakarma, S., & Pujahari, A. (2017). *Evaluation of machine learning models for employee churn prediction*. Paper presented at the 2017 International Conference on Inventive Computing and Informatics (ICICI).
- Sivakumar, S., Videla, L. S., Kumar, T. R., Nagaraj, J., Itnal, S., & Haritha, D. (2020). *Review on word2vec word embedding neural net*. Paper presented at the 2020 international conference on smart electronics and communication (ICOSEC).
- Srinivas, A. C. M. V., Satyanarayana, C., Divakar, C., & Sirisha, K. P. (2021). *Sentiment analysis using neural network and LSTM*. Paper presented at the IOP conference series: materials science and engineering.
- Sugumaran, P., & Uma, A. (2022). Real-time twitter data analytics of mental illness in COVID-19: sentiment analysis using deep neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1), 560-567.
- Trupthi, M., Pabboju, S., & Narasimha, G. (2017). *Sentiment analysis on twitter using streaming API*. Paper presented at the 2017 IEEE 7th International Advance Computing Conference (IACC).
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.
- Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems (TMIS)*, 9(2), 1-29.

Corresponding author: Alexander Ntino.

✉ alexntino@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi.

© 2023. Faculty of Technology Education. ATBU Bauchi. All rights reserved