International Journal of Engineering and Management Research
e-ISSN: 2250-0758 | p-ISSN: 2394-6962
https://ijemr.vandanapublications.com

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

Sentiment Analysis of Twitter Data Using Machine Learning Techniques

Mantasha Khan[1] and Ankita Srivastava[2]
[1]Student, Department of Computer Science & Engineering, Integral University, INDIA
[2]Assistant Professor, Department of Computer Science & Engineering, Integral University, INDIA

[1]Corresponding Author: mantashakhan900.0@gmail.com

Received: 22-01-2024 Revised: 11-2-2024 Accepted: 28-02-2024

ABSTRACT

In the age of social media, it is more convenient for individuals to articulate their thoughts and emotions. Each day, they disseminate their perspectives and notions on various social media platforms about ongoing global events. On controversial issues, one can find a consensus of public feeling, whether positive or negative. Twitter functions as a demonstration of a social media platform where individuals participate in discussions about their perspectives. Twitter sentiment analysis examines the overall feeling or emotion expressed in tweets. It employs machine learning and natural language processing techniques to automatically categorize tweets as good, negative, or neutral depending on their content. It may be used for single tweets or a bigger dataset relating to a specific topic or event. Through the identification of these sentiments, machine learning endows us with an advantageous position in the analysis and prediction of said sentiments. Distinct machine learning models are utilized in this paper to scrutinize sentiments within Twitter data. The proposed system offers a comprehensive evaluation of the performance of various machine learning algorithms, including Vader, XGBoost with CountVectorizer, XGBoost with Gensim, Random Forest with CountVectorizer, Random Forest with Gensim, Single LSTM, and Bidirectional LTSM and Bidirectional LTSM gives highest accuracy of .73.

Keywords-- Crisis Management, LSTM, Sentimental Analysis, Tokenization, Vader

I. INTRODUCTION

Twitter has emerged as a prominent platform for discussion of intense emotions, making it a valuable source of information for analyzing sentiments. Sentiment analysis is the technique of examining text to detect its underlying emotional tone. With the rise of social media platforms like Twitter, analysis of sentiment has become an essential tool for businesses, associations, and governments seeking to comprehend public opinion and form well-informed perspectives [11]. Natural Language Processing (NLP) methods are extensively employed for sentiment analysis as they enable machines to comprehend and interpret human language [12]. NLP techniques can analyze tweets in real time, identify the sentiment conveyed in tweets, and provide insights into prevailing trends and patterns in public sentiment [13]. Machine learning algorithms, which fall under the umbrella of NLP, can acquire knowledge from vast datasets and accurately predict the sentiment of new tweets. In this investigation, we aim to assess the efficacy of ML systems in conducting sentiment analysis on Twitter using NLP methodologies. To classify tweets as favorable, negative, or neutral., we will utilize a dataset that includes tweets from the opening day of the "FIFA World Cup 2022", held in Qatar. This dataset encompasses information such as the date of creation, number of likes, tweet source, tweet content, and sentiment. We will preprocess this data to eliminate any noise and subsequently use machine learning methods such as Vader, XGBoost, Random Forest, and LSTM (Long Short-Term Memory). To improve accuracy, we use count vectorizers and genism in our models. Machines cannot interpret letters or words. When dealing with text data, we must represent it numerically so that the machine can interpret it. Count vectorizer is a method for translating text to numerical data. Gensim is an open-source Python package for NLP. The Gensim package this allows us to create word embeddings by instruction word2vec classifiers on a particular corpus using either the CBOW or skip-gram approaches. The effectiveness of these algorithms will be evaluated based on several criteria, including F1 score, accuracy, recall, and precision. By employing sentiment analysis, organizations can make well-informed decisions regarding real-time monitoring, audience engagement, brand impression, fan experience enhancement, and crisis management. Twitter sentiment analysis is crucial because it helps businesses understand customer feedback and find areas where their products or services may be Improved [14]. Sentiment analysis may help businesses track their company reputation online and react promptly to unfavorable comments or reviews [15]. Sentiment research may help political campaigns better grasp public sentiment and modify their messaging accordingly [16]. In the case of a crisis, sentiment analysis may assist companies in monitoring social media and news channels for adverse emotions and responding accordingly. Marketers may use sentiment analysis

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

to comprehend customer habits and tastes, as well as build customized advertising campaigns [17].

II. LITERATURE REVIEW

Pak and Paroubek (2010) [7] proposed an approach for categorizing tweets as objective, good, and negative. They gathered tweets using the Twitter API to create a Twitter corpus. The tweets are automatically labeled with emoticons. They developed a sentiment classifier utilizing a Naive Bayes algorithm using features such as Ngrams and POS tags. The training set they used was less effective since it only comprised tweets with emoticons.

N Bahrawi (2019) [3] This study uses the Random Forest technique to analyze sentiments using Twitter data sources. They will assess the results of the assessment of the method used in this study. The margin of error of observations in this investigation was roughly 75%.

G. Shobana et al.(2019) [9] The research examines celebrity IDs (@realdonaldtrump) or hashtags (#IPL2018) to gain insight into the attitudes of individuals on every occasion when the individual tweets or acts in certain situations. The suggested method would assess people's sentiments utilizing Python, Twitter API, and Text Blob (a text processing library). As a consequence, it allows for a more accurate examination of the post.

D. Ramana Kumar et al.(2020) [8] In this work, the Bi-LSTM generated three sorts of outcomes: favorable, adverse, and zero to verify the TSA for the Sanders collection. In comparison to current approaches including SVM and Neural Networks, the new Bi-LSTM approach obtained 90.04 percent efficiency, 88.12 percent precision, 92.31 percent recall, and 90.17 percent F-Measure.

S. Jacob et al. (2021) [4] In this article, the researchers applied a machine learning-based clustering technique. Tests were performed in a qualifying and test collection comprised of enormous amounts of tweets from data with over one lakh results, demonstrating efforts to detect whether a tweet is either positive or negative.

G. Ravi Kumar et al.(2021) [5] This paper attempted to employ three distinct machine learning approaches to conduct an estimate assessment. The critical assessment is to determine the intensity of the material and categorize it as good, bad, or zero sentiment in the tweets. As a result, the primary purpose of this investigation work is to conduct estimation investigations utilizing machine learning (ML) approaches for sentiment analysis. The ML experiments are carried out while employing a US airline Twitter informative index obtained from the Kaggle. The effectiveness of all three ML clusters techniques, including decision trees, SVMs, and neural networks, are examined, with an emphasis on reliable information. The neural network-based technique achieved remarkable accuracy (75.99%).

Cihan ÇILGIN et al.(2022) [1] The researchers collected the tweets with hashtags like '#covid19', '#Covid', '#pandemic', '#social-distancing', '#socialdistance', '#covid-19', '#corona-virius', '#coronavirus', '#Chinesevirus', '#Chinese-virus' . Between January 1 and July 1, 2020, they gathered tweets from Twitter an overall of 60,243,040 tweets. In this study, they employed VADER to categorize the emotion conveyed in Twitter data connected to COVID-19, and the overall scores of the following tweets were separated into five groups. Furthermore, in the research, Word cloud was utilized to depict the most often gathered text data every month, whereas N-grams were used to comprehend tweets and their meaning. Although there were more unfavorable tweets regarding COVID-19 during the earlier phases of the epidemic, users posted more favorable tweets later on.

Lal Khan et al.(2022) [6] In this paper, researchers have used the CNN-LSTM Model layout using conventional machine learning algorithms. They provide an innovative deep learning framework for English dialect SA as well as Roman Urdu, which consists of two distinct levels: an LSTM for ongoing dependency maintenance and a single-layer CNN algorithm for geographical feature extraction. To acquire the ultimate classification, CNN and LSTM feature maps are input into many machine learning models. Several word embedding algorithms lend credence to this idea. Comprehensive evaluations on four datasets show that the suggested approach is extremely efficient in English text and Roman Persian categorization of sentiment, with success rates around 0.904, 0.841, 0.740, as well as 0.748 vs MDPI, RUSA, RUSA-19, and UCL records, respectively. The outcomes indicate that the classification algorithm using SVM and the Word2Vec CBOW framework are better alternatives for Roman Urdu sentiment evaluation. In contrast, BERT embedding of words, two-layer LSTM, and SVM as categorical works are more efficient alternatives for sentiment evaluation in English. The proposed model surpasses other well-known sophisticated methods on related databases, increasing performance by more than 5%.

Richa Dhanta et al.(2023) [2] In this article, they examined the dataset from Twitter for sentiment whether they are favorable, unfavorable, or zero. They used a dataset of tweets collected from different sources, which were then preprocessed to remove noise and improper information . To divide tweets as favorable, unfavorable, or neutral, several machine learning techniques were used, such as logistic regression and Naive Bayesian. The efficiency of these methods is also assessed in the study using a different number of criteria, including F1 score, accuracy, recall, and precision. The results indicate that

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

machine learning methods are effective in analyzing sentiment on Twitter, with Naive Bayes providing the best efficiency.

III. METHODOLOGY

The sentiment analysis technique involves numerous phases. The first step is to gather data and execute preprocessing. This dataset was obtained from the Kaggle repository and comprises tweets from the inaugural day of the FIFA World Cup 2022, which took place in Qatar. In the pre-processing stage, we preprocess the dataset by cleaning the tweets, eliminating usernames, URLs, stopwords, and so on, and then apply lemmatization techniques to standardize words to ensure consistency in sentiment analysis. To improve the accuracy of our model, we integrate CountVectorizer and Gensim Word2Vec Model with our machine learning methods. This is accomplished through the use of feature extraction and feature selection techniques. These techniques are used to reduce the amount of input variables, prevent overfitting, reduce computing complexity and training time, and increase model accuracy. Feature extraction methods include vectorization and word embedding. Finally, machine learning algorithms are employed to identify text as positive, negative, or neutral depending on sentiment polarity. Machine learning algorithms categorize feelings based on training and test datasets. The machine learning models we use are Vader, XGBoost, Random Forest, and LSTM (Long Short-Term Memory), are presented here. (Figure 1)

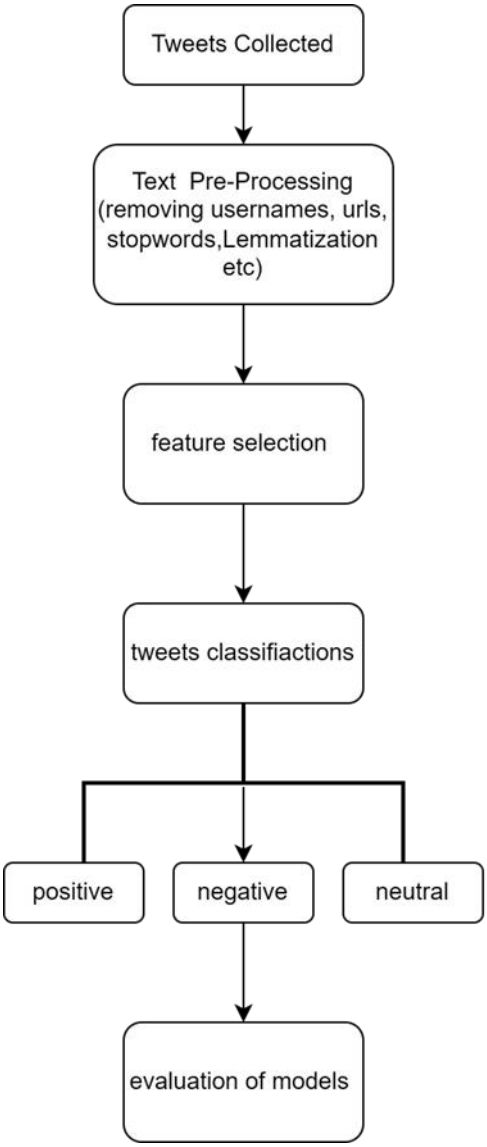

Figure 1: Methodology

A. Data Collection

This dataset was acquired from the Kaggle repository and includes tweets from the opening day of the "FIFA World Cup 2022" held in Qatar. ("https://www.kaggle.com/code/aks777sp/fifa-world-cup-day-1-tweets"). In this dataset, we have 22524 rows and 6 columns. The dataset (Figure 2) contains information such as the unnamed, date created, the number of likes, the source of the tweet, the tweet itself, and the sentiment.

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

	Unnamed: 0	Date Created	Number of Likes	Source of Tweet	Tweet	Sentiment
0	0	2022-11-20 23:59:21+00:00	4	Twitter Web App	What are we drinking today @TucanTribe \n@MadB...	neutral
1	1	2022-11-20 23:59:01+00:00	3	Twitter for iPhone	Amazing @CanadaSoccerEN #WorldCup2022 launch ...	positive
2	2	2022-11-20 23:58:41+00:00	1	Twitter for iPhone	Worth reading while watching #WorldCup2022 htt...	positive
3	3	2022-11-20 23:58:33+00:00	1	Twitter Web App	Golden Maknae shinning bright\n\nhttps://t.co/...	positive
4	4	2022-11-20 23:58:28+00:00	0	Twitter for Android	If the BBC cares so much about human rights, h...	negative

Figure 2: The Dataset

B. Pre-Processing

For the model to be more accurate, we must preprocess the data. To preprocess the tweets collected, we first remove the usernames, URLs, stopwords, and so on. After we have removed all of the usernames, URLs, and stopwords, we tokenize the text and utilize the lemmatizer approach to reduce the term to its root form.

Table 1: Tweets along with pre-processed tweets

Tweets	Pre-Processed Tweets
'England â\x9c\x85 @England #WorldCup2022 https://t.co/Zy3uPDRfWI'	'england worldcup'
'USA v Wales tomorrow. A must win? #WorldCup2022'	'usa v wale tomorrow must win worldcup'
'I'm going for @England to win 3-0 tomorrow #ENGIRN #WorldCup2022'	'im going win tomorrow engirn worldcup'
'The tunnel footage of the Qatar players coming out today was smart!!! #WorldCup2022 #QATECU https://t.co/LkLROGJ1Go'	'tunnel footage qatar player coming today smart worldcup qatecu'
'@anonewsco , @KromSec broadcast Iran Protests in #WorldCup2022 #FIFAWorldCup \n\n#MahsaAmini\n#Anonymous \n#OpIran'	'broadcast iran protest worldcup fifaworldcup mahsaamini anonymous opiran'

C. Machine Learning Models

a. VADER

In the nltk, sentiment, Python library provides a module called VADER, which was designed primarily to handle text generated in social networking settings, while it can also handle language from other contexts. VADER can determine the polarity of sentiments (positive or negative) in a particular chunk of text when the data is processed unlabeled. To identify the overall sentiment of a corpus of text, VADER consults a lexicon of sentiment-related terms.

Here's an example of how the lexicon is organized, with each word assigned a valence rating(In Figure 3)

Word	Sentiment rating
Tragedy	-3.4
Rejoiced	2
Insane	-1.7
Disaster	-3.1
Great	3.1

Figure 3: Lexicon with valence rating

b. XGBoost

XGBoost is a machine learning method that falls under the ensemble learning category, especially the gradient boosting framework. It uses decision trees as base learners and regularization approaches to improve model generalization. XGBoost, known for its computing speed, feature significance analysis, and management of missing values, is commonly used for applications like regression, classification, and ranking.

XGBoost, or eXtreme Gradient Boosting, is a machine learning technique classified as ensemble learning. It is used for supervised learning problems like regression and classification. XGBoost creates a predictive model by iteratively merging the predictions of numerous independent models, most often decision trees.

c. Random Forest

A Random Forest is similar to a collaborative decision-making team in machine learning. It integrates the opinions of several "trees" (individual models) to improve predictions, resulting in a stronger and better-performing model. The Random Forest Algorithm's extensive appeal originates from its user-friendliness and versatility, which allow it to efficiently handle both classification and regression issues. The algorithm's strength is its ability to handle complicated datasets while minimizing overfitting, making it a useful tool for a variety of prediction tasks in machine learning.

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

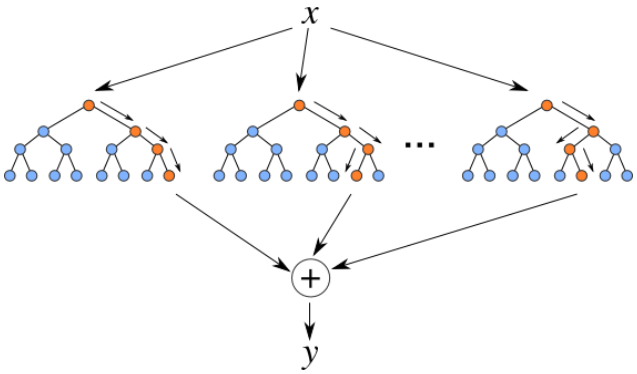

Figure 4: Random Forest

d. LSTM (Long Short-Term Memory)

(LSTM (Long Short-Term Memory) is a form of RNN (Recurrent Neural Network) that can maintain long-term dependencies in sequential input. LSTMs can process and evaluate sequential data, including time series, text, and voice. The functioning of an LSTM can be shown in(Figure.5***)***

LSTMs have demonstrated exceptional success in sentiment analysis on tweets because of their capacity to capture contextual information, manage varied input durations, and model sophisticated language patterns. They are an effective method for extracting sentiment-related information from brief and informal writing, resulting in a better knowledge of public opinion and sentiment patterns on social media platforms**.**

Bidirectional LSTM, often known as BiLSTM, refers to a model of sequences that has two LSTM layers, one for forward processing and another for backward processing. It is typically used for NLP-related activities. The idea behind this method is that by analyzing input in both ways, the model may better grasp the link between sequences.

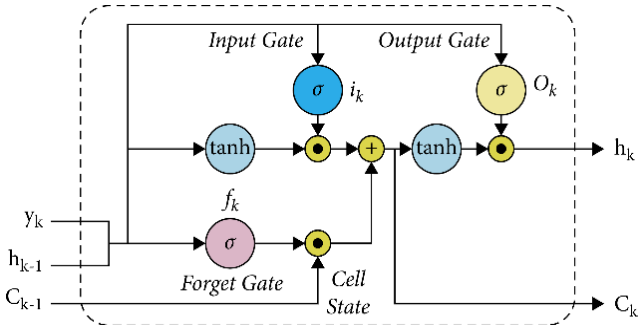

Figure 5: LSTM architecture

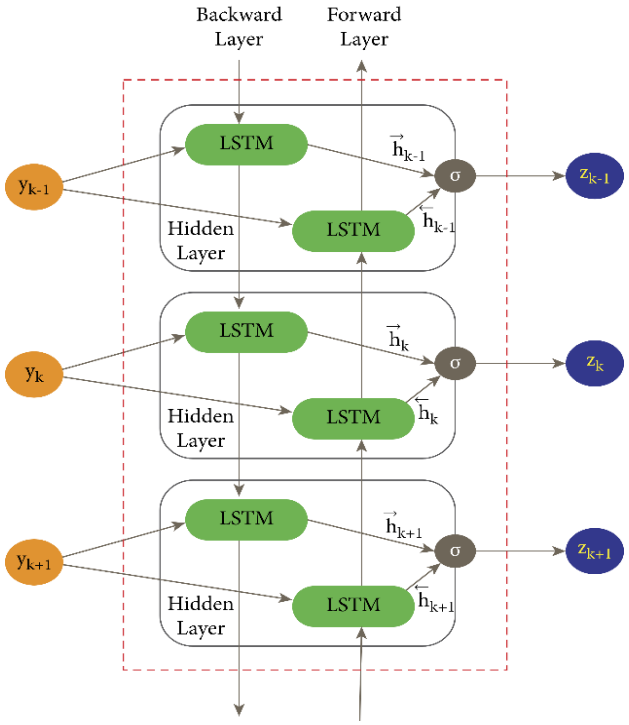

Figure 6: Bidirectional LSTM architecture

IV. RESULTS AND DISCUSSIONS

Sentiment analysis is assessed through the consideration of metrics such as accuracy, precision, recall, and F1-Score (represented by equations 1, 2, 3, and 4). The sentiment distribution, following the preprocessing of the dataset, is visualized in Figure 7. In addition, Figure 8 showcases the Top 10 Sources of Tweet Count. Moreover, Figure 9 presents a Word Cloud depicting sentiments. Lastly, Figure 10 exhibits the time series sentiment trends on the initial day of the FIFA World Cup 2020 in Qatar. Emotion labels are resampled and tallied at various time intervals.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

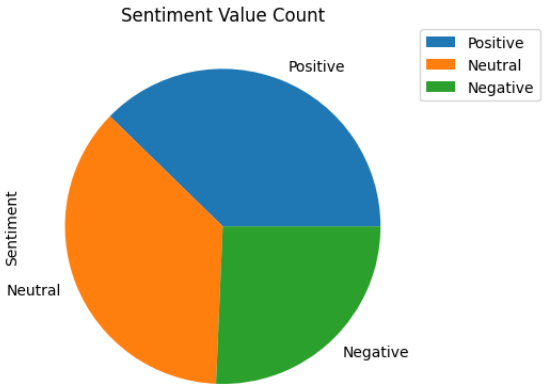

Figure 7: Distribution of sentiment

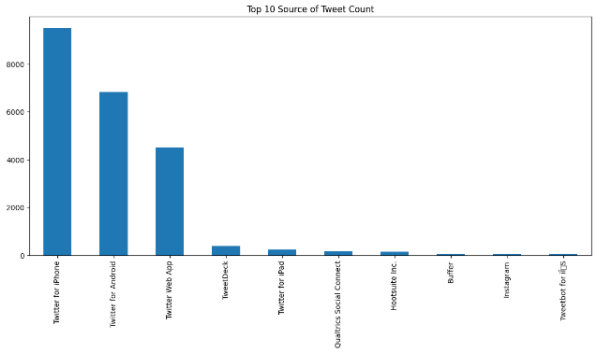

Figure 8: Top 10 Sources of Tweet Count

Figure 9: Word Cloud

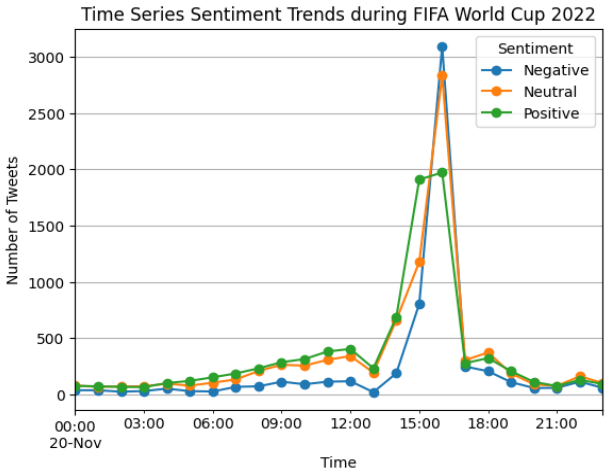

Figure 10: Time series Sentiment Trends

Table 2: Performance Analysis of Machine Learning Models for Sentiment analysis

Models	Accuracy
Vader	.57
XGBoost with CountVectorizer	.70
XGBoost with Gensim	.59
Random Forest with CountVectorizer	.69
Random Forest with Gensim	.68
Single LSTM	.71
Bidirectional LTSM	.73

Many machine learning methodologies, such as Vader, XGBoost with CountVectorizer, XGBoost with Gensim, Random Forest with CountVectorizer, Random Forest with Gensim, Single LSTM, and Bidirectional LTSM, are employed to conduct sentimental analysis. Amongst all of these machine learning models that we have tested, it is observed that Bidirectional LTSM yields the most superior outcomes, with an accuracy of 0.73. When compared to the other machine learning models, Bidirectional LTSM exhibits the greatest performance. The performance analysis of machine learning models for sentiment analysis is illustrated in Table 2. In addition, Figure 11 presents the confusion matrix of the bidirectional LSTM model, which possesses the highest degree of accuracy. The F1-score, recall, and precision all amount to 0.73. Figure 13 demonstrates the evaluation metrics of our machine learning approaches.

Peer Reviewed & Refereed Journal
Volume-14, Issue-1 (February 2024)
https://doi.org/10.5281/zenodo.10791485

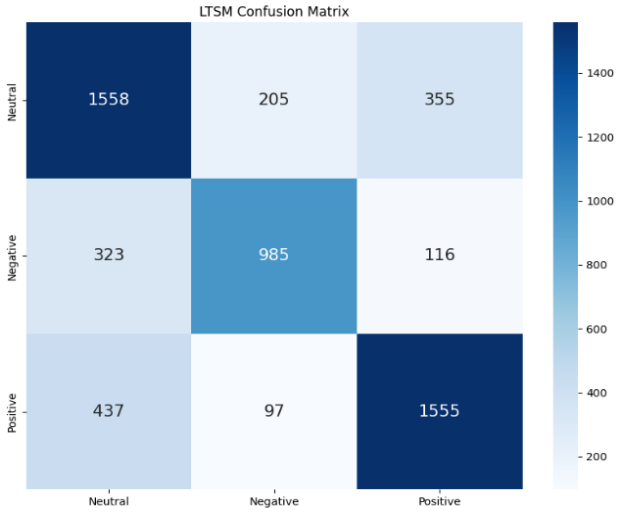

Figure 11: Confusion Matrix of Bi-LSTM Model

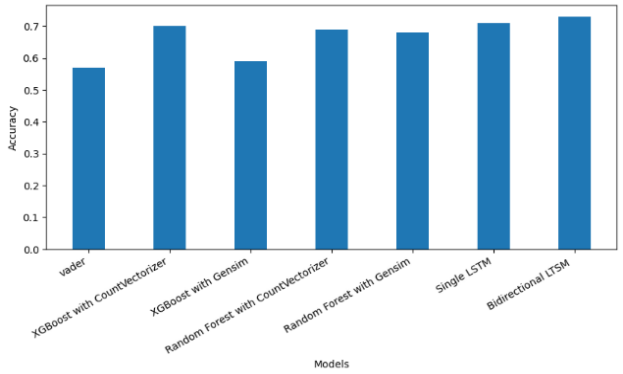

Figure 12: Accuracy scores of Machine learning Approaches

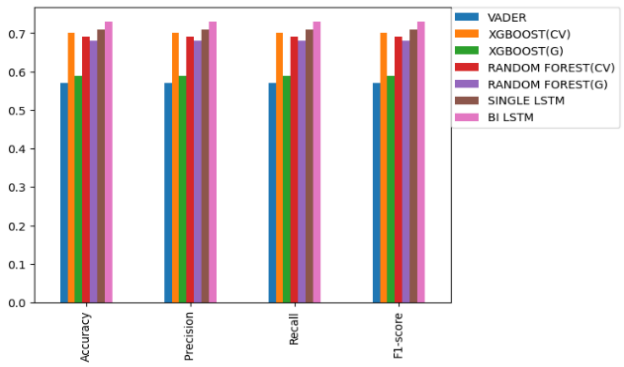

Figure 13: Graphical representation of evaluation performance of algorithms

V. APPLICATIONS

1. **Real-time monitoring:** It allows stakeholders to address negative opinions quickly while capitalizing on good sentiment.
2. **Audience Engagement:** Marketing teams may adapt their efforts based on sentiment analysis data to better connect with their target audience.
3. **Brand impression:** Sponsors can assess their brand's impression among fans and alter strategy appropriately.
4. **Fan Experience Enhancement:** Event organizers may use sentiment analysis to identify areas for improvement and then improve the entire fan experience, resulting in maximum satisfaction.
5. **Crisis Management:** Quickly identify possible controversies or unfavorable situations and take corrective steps to limit their impact.

VI. CONCLUSION

This paper examines the utilization of various machine learning methodologies, including Vader, XGBoost with CountVectorizer, XGBoost with Gensim, Random Forest with CountVectorizer, Random Forest with Gensim, Single LSTM, and Bidirectional LTSM, for performing sentiment analysis on Twitter. Among all these machine learning models that have been tested, it has been observed that Bidirectional LTSM produces the most superior results, achieving an accuracy of 0.73. When compared to the other machine learning models, Bidirectional LTSM demonstrates the highest performance. Initially, the data from the Kaggle dataset "fifa_world_cup_2022_tweets", which was stored in a CSV file, will be loaded. This dataset comprises tweets related to the opening day of the FIFA World Cup held in Qatar. Subsequently, the dataset will be pre-processed by performing tasks such as removing usernames, URLs, stopwords, lemmatization, and tokenization. Furthermore, visualizations will be created to gain insights, including sentiment distribution plots, word clouds depicting positive and negative words, and time series sentiment trends during the event. Additionally, sentiment scores will be calculated for each tweet by leveraging these models, encompassing dimensions of positivity, negativity, and neutrality. Lastly, a comparative study among these models will be conducted. In the future, the Neural Network model shows promise and has the potential to outperform the other models in terms of accuracy if it is fine-tuned.

REFERENCES

[1] Çilgin, C., Baş, M., Bilgehan, H. & Ünal, C. (2022). Twitter sentiment analysis during covid-19 outbreak with VADER. *AJIT-e: Academic Journal of Information Technology, 13*(49), 72–89. https://doi.org/10.5824/ajite.2022.02.001.x.

[2] Dhanta, R., Sharma, H., Kumar, V. & Singh, H. O. (2023). Twitter sentimental analysis using machine learning. *International Journal of Communication and Information Technology, 4*(1), 71–83. DOI: 10.33545/2707661x.2023.v4.i1a.63.

[3] Bahrawi, N. (2019). Sentiment analysis using random forest algorithm-online social media based. *Journal of Information Technology and Its Utilization, 2*(2), 29. https://doi.org/10.30818/jitu.2.2.2695.

[4] Jacob, S. S. & Vijayakumar, R. (2021). Sentimental analysis over twitter data using clustering based machine learning algorithm. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-02771-9.

[5] Ravi Kumar, G., Venkata Sheshanna, K. & Anjan Babu, G. (2021). Sentiment analysis for airline tweets utilizing machine learning techniques. In: *EAI/Springer Innovations in Communication and Computing*, pp. 791–799. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-49795-8_75.

[6] Khan L, Amjad A, Afaq KM & Chang H-T. (2022). Deep sentiment analysis using CNN-LSTM architecture of english and roman urdu text shared in social media. *Applied Sciences, 12*(5), 2694. https://doi.org/10.3390/app12052694.

[7] Alexander Pak & Patrick Paroubek. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta*. European Language Resources Association.

[8] Kumar, D. & Rao, S. (2020). *A sentiment analysis of twitter data using bi-directional long short term memory*. DOI: 10.1007/978-3-030-30271-9_16.

[9] Shobana, G., Vigneshwara, B. & Maniraj Sai, A. (2019). Twitter sentimental analysis. *International Journal of Recent Technology and Engineering, 7*(4), 343–346. https://doi.org/10.46501/ijmtst061266.

[10] Dashrath Mahto, Subhash Chandra Yadav & Gotam Singh Lalotra. (2022). Sentiment prediction of textual data using hybrid convbidirectional-lstm model. *Mobile Information Systems*. https://doi.org/10.1155/2022/1068554.

[11] I. Guellil & K. Boukhalfa. (2015). Social big data mining: A survey focused on opinion mining and sentiments analysis. *12th International Symposium on Programming and Systems (ISPS), Algiers, Algeria*, pp. 1-10. DOI: 10.1109/ISPS.2015.7244976.

[12] A. Świetlicka, D. Haczyk & M. Haczyk. (2023). Graph neural networks for natural language processing in human-robot interaction. *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland*, pp. 89-94. DOI: 10.23919/SPA59660.2023.10274451.

[13] K. S. Madhu, B. C. Reddy, C. Damarukanadhan, M. Polireddy & N. Ravinder. (2021). Real time sentimental analysis on twitter. *6th International Conference on Inventive Computation Technologies, Coimbatore, India*, pp. 1030-1034. DOI: 10.1109/ICICT50816.2021.9358772.

[14] N. Deepa, J. S. Priya & T. Devi. (2023). Sentimental analysis recognition in customer review using Novel-CNN. *International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India*, pp. 1-4. doi: 10.1109/ICCCI56745.2023.10128627.

[15] Y. E. Cakra & B. Distiawan Trisedya. (2015). Stock price prediction using linear regression based on sentiment analysis. *International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia*, pp. 147-154. DOI: 10.1109/ICACSIS.2015.7415179.

[16] P. Khurana Batra, A. Saxena, Shruti & C. Goel. (2020). Election result prediction using twitter sentiments analysis. *Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India*, pp. 182-185. DOI: 10.1109/PDGC50313.2020.9315789.

[17] A. Z. Adamov & E. Adali. (2016). Opinion mining and Sentiment Analysis for contextual online-advertisement. *IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan*, pp. 1-3. DOI: 10.1109/ICAICT.2016.7991682.

[18] Malde, Ravi. (2020). A short introduction to VADER. *Towards Data Science*. Available at: https://towardsdatascience.com/an-short-introduction-to-vader-3f3860208d53.

[19] Schott, Madison. (2019). Random forest algorithm for machine learning. *Medium*. https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb.