iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53121

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Tweet Sentiment & Emotion Analysis

Jeevan Nair PK[1], Sudharsan V[2], Kiran K Iyer[3], Prof. Rosemary Varghese[4]
Department of Computer Science & Engineering, Adi Shankara Institute of Engineering and Technology, Kalady, Kerala, India

Abstract: ***Tweet Sentiment & Emotion Analysis using Bi-LSTM in RNN. Twitter has developed into a useful medium for sharing ideas, attitudes, and feelings. Applications like opinion mining, market research, and social trend analysis all depend on the sentiment and emotion of tweets. The Bi-LSTM architecture in RNN (Recurrent Neural Networks) is used in this study to present an advanced method for sentiment and emotion analysis on Twitter. By utilising machine learning techniques, the objective is to increase the precision and efficacy of mood and emotion analysis. The study emphasises both conventional text analysis and real-time data analysis. Organisations and governmental bodies may continuously monitor sentiment and emotional patterns on Twitter thanks to real-time analysis, which enables them to react quickly to emerging problems or crises. In a typical text analysis, historical tweet data is examined to learn more about user viewpoints, emotional patterns, and sentiment distributions. The Bi-LSTM architecture is used because it can effectively capture the context and sequential dependencies found in tweets. To ensure consistent analysis, the system gathers real-time tweets and conducts the preprocessing stages. Monitoring sentiment changes, emotional responses, and new trends on Twitter are all made possible by this study. The goal of the research is to improve the precision and efficacy of sentiment and emotion analysis on Twitter by utilising machine learning techniques, real-time data analysis, and standard text analysis. The findings and conclusions will aid in understanding public mood and feelings in the digital age.***

Keywords: Twitter sentiment analysis, emotion analysis, Bi-LSTM, RNN, Natural language processing, real-time data analysis

I. INTRODUCTION

Twitter sentiment and emotion analysis is determining the sentiment and emotion of tweets using natural language processing (NLP) and machine learning techniques. Sentiment analysis is the process of recognising and categorising positive, negative, or neutral opinions conveyed in text. Emotion analysis, on the other hand, entails determining the underlying emotion underlying a tweet.

The Bi-directional Long Short-Term Memory (Bi-LSTM) model is a prominent machine learning technique for sentiment and emotion analysis. The Bi-LSTM model is a recurrent neural network (RNN) that can process sequential data like text.

The Bi-LSTM model is made up of many layers of LSTM cells. Each LSTM cell is in charge of remembering the prior state as well as determining the current state.

The Bi-LSTM model is referred to as "bi-directional" since it processes the input sequence both forward and backward. As a result, the model is able to capture the context of the full input sequence.

The Bi-LSTM model can be trained on a large dataset of labelled tweets in the context of Twitter sentiment and emotion research. Based on the sentiment conveyed, the labelled tweets can be classified as good, negative, or neutral. The dataset can also be labelled according to the underlying emotion expressed in the tweet, such as anger, joy, sadness, or fear.

Once trained, the Bi-LSTM model can be used to predict the sentiment and emotion of new tweets. The model takes a tweet's word sequence as input and generates a probability distribution across the possible sentiment or emotion categories.

The Bi-LSTM model modifies the weights of its neurons during training to minimise the error between anticipated and real sentiment or emotion labels. This is known as backpropagation, and it is an important part of neural network training.

Other machine learning methods, such as convolutional neural networks (CNNs) and support vector machines (SVMs), can be employed in addition to the Bi-LSTM model for Twitter sentiment and emotion analysis. However, the Bi-LSTM model has been shown to be particularly effective for analyzing sequential data such as text.

Finally, Twitter sentiment and emotion analysis using the Bi-LSTM model is an effective method for analysing the sentiment and emotion communicated in tweets. With the growing usage of social media and the massive volumes of data created by these platforms, sentiment and emotion analysis can provide significant insights into people's opinions and emotions on a variety of topics and concerns.

© IJRASET: All Rights are Reserved |

II. EXISTING SYSTEM

The current sentiment and emotion analysis system for Twitter makes use of conventional machine learning techniques including feature engineering and classification algorithms. To infer sentiment and mood from tweets, it may also use rule-based systems or lexicon-based techniques. The abundance of sarcasm, figurative language, and changing linguistic trends on Twitter are just a few of the problems these algorithms frequently run into when dealing with unstructured text data.

The lexicon-based strategy was a popular approach. These systems used sentiment lexicons created specifically for Twitter data. Lexicons contained annotated words and phrases with sentiment scores. Sentiment analysis algorithms would compute a tweet's overall sentiment by summing the sentiment scores of its constituent terms. While this method was straightforward and easy to use, it struggled with slang, neologisms, and the ever-changing nature of Twitter language.

Another method required the use of machine learning methods like Support Vector Machines (SVM) or Naive Bayes. These methods need feature engineering, which involves extracting multiple features from the text, including word frequencies, n-grams, part-of-speech tags, and syntactic patterns. These features were then used to train a model that could categorise tweets based on their emotion. However, these models frequently struggled to capture contextual information and nuances of Twitter language.

The usage of sentiment-specific features was a popular strategy for Twitter sentiment analysis. Positive and negative emoticons, prolonged words (e.g., "loooove"), repeated letters (e.g., "happyyyy"), and capitalization were among these characteristics. These features attempted to capture the sentiment portrayed in the text using specific patterns seen on Twitter. This technique, however, was significantly reliant on handcrafted features and lacked the ability to learn complex patterns automatically.

Prior to the introduction of RNNs, current algorithms for Twitter sentiment analysis struggled to capture the contextual information, sarcasm, irony, and other nuances inherent in Twitter data. RNNs revolutionised sentiment analysis by modelling long-term dependencies and capturing the context in which words and phrases arise, thanks to their capacity to capture sequential dependencies in text. This resulted in considerable increases in sentiment analysis performance, notably in capturing Twitter data's distinctive properties.

The following are some of the system's drawbacks:

- Inability to understand the context and sequential dependencies in tweets.
- Limited accuracy as a result of the use of manually created features and rule-based techniques.
- Challenges in adjusting to Twitter's changing linguistic trends and real-time data processing.

III. OBJECTIVES

The Bi-LSTM model is used for Twitter sentiment and emotion analysis in order to accurately predict the sentiment and emotion represented in tweets. This technique has the ability to provide useful insights into people's opinions and emotions about a variety of issues and concerns.

The precise objectives and goals related with this analysis are as follows:

1) Creating a robust and accurate model: By leveraging the Bi-LSTM model, the objective is to develop a sentiment and emotion analysis model that is both reliable and precise. The Bi-LSTM model, with its ability to capture contextual information and long-term dependencies in text, can effectively handle the challenges posed by the unique characteristics of Twitter data. It provides a powerful framework to learn and understand the sentiment and emotional tone expressed in tweets.
2) Enhancing analysis accuracy with context: The addition of bi-directional processing in the Bi-LSTM model allows for the incorporation of the full context of tweets. This is crucial for sentiment and emotion analysis as it enables a more comprehensive understanding of the text. By considering the complete tweet, including its preceding and succeeding words, the model can better capture the sentiment nuances, sarcasm, and other contextual elements that impact the overall sentiment and emotion expressed.
3) Providing actionable information for entities: By analyzing the sentiment and emotion of tweets related to a company's brand, products, or services, valuable insights can be generated. These insights can inform corporations, organizations, and individuals about the public's perception and sentiment towards their offerings. It enables them to understand customer satisfaction, identify areas for improvement, and make informed decisions to enhance their brand reputation and customer experience.

© IJRASET: All Rights are Reserved |

4) Understanding social media trends and public opinion: Analyzing large volumes of Twitter data allows for a deeper understanding of social media trends and public sentiment. By mining and analyzing this vast amount of user-generated content, patterns and shifts in sentiment can be identified. This information can be valuable for market research, public opinion analysis, and tracking the response to specific events, campaigns, or policies.
5) Real-time sentiment and mood monitoring: The application of the Bi-LSTM model enables real-time sentiment and mood monitoring on Twitter. By continuously analyzing tweets as they are posted, changes in public opinion and sentiment can be detected promptly. This provides an opportunity for organizations and individuals to stay updated on the evolving sentiment and respond in a timely and appropriate manner. It facilitates proactive engagement, crisis management, and the ability to address concerns or issues swiftly.

IV. PROPOSED SYSTEM

The following steps would be involved in the proposed system for Twitter sentiment and emotion analysis using Bi-LSTM in RNN:

1) *Data Collection:* The initial stage would be to acquire a big collection of tweets labelled with sentiment and emotion categories.
2) *Pre-processing:* Any extraneous information, such as URLs, user mentions, and hashtags, would be removed from the obtained dataset. In addition, the text would be cleaned by removing stop words, punctuation, and converting all text to lowercase.
3) *Feature Extraction:* Using techniques such as word embedding, the pre-processed dataset would be utilised to extract features. Word embedding is the process of expressing each word in a tweet as a high-dimensional vector that captures the semantic meaning of the word.
4) *Model Training:* The pre-processed and feature-extracted dataset would be used to train the Bi-LSTM model. During training, the model's weights would be changed to minimise the difference between anticipated and actual sentiment or mood labels.
5) *Model Evaluation:* The trained model will be tested on a separate dataset to see how well it predicts sentiment and emotion. The model's performance would be measured using metrics like accuracy, precision, and recall.
6) *Deployment:* Once trained and assessed, the model will be used to analyse new tweets in real time. Tweets can be pre-processed and features retrieved using the same techniques as during training. The Bi-LSTM model can then be used to forecast the tweet's sentiment and emotion.

Overall, the proposed Twitter sentiment and emotion analysis system based on Bi-LSTM in RNN can provide accurate and important insights into the sentiment and emotion represented in tweets. This can assist businesses, organisations, and individuals in making educated decisions based on public opinion and so improving their products and services.

Algorithm 1 Twitter Authentication

```
for Valid Twitter Account do
    Consumer-key=" "
    Consumer-secret=" "
    Access-token=" "
     AccessTokenSecret=" "
    if (flag=TRUE) then
        authenticate with OAuthHandler(Consumer-
            key, Consumer-Secrete-Key)
        set the authentication with access-key and
        access- secrete-key
        Then enable to a twitter-stream
    end if
     Authentication failure
end for
```

V. SOFTWARE SETUP

The proposed system will require the following software setup:

1) Python is the programming language used to implement the system.
2) The well-known machine learning framework TensorFlow is used to create and train neural networks.
3) Tweepy is a Python package that allows users to collect real-time tweets by using the Twitter API.
4) Python text-processing and sentiment analysis library called TextBlob is also used.
5) Django is the framework used for the front end.

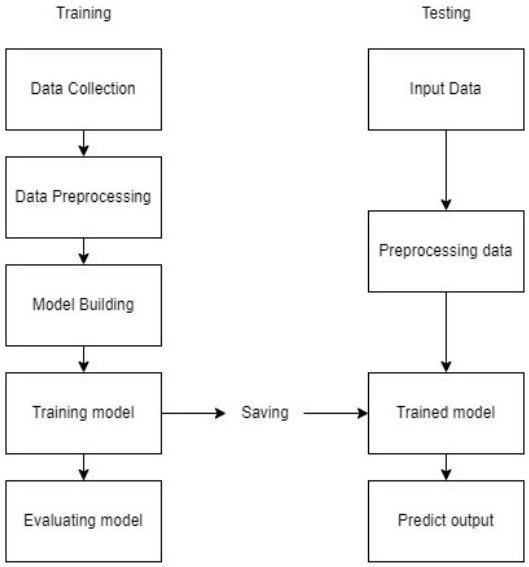


Fig: System Architecture

VI. RESULT

The system's performance is promising, as evidenced by the accuracy of 87.10% obtained in the project utilising Bi-LSTM in RNN. The percentage of accurately identified sentiments and emotions in the dataset is represented by the accuracy metric. Being able to anticipate events with such high accuracy shows that the Bi-LSTM model successfully captures the sequential dependencies and context of tweets.

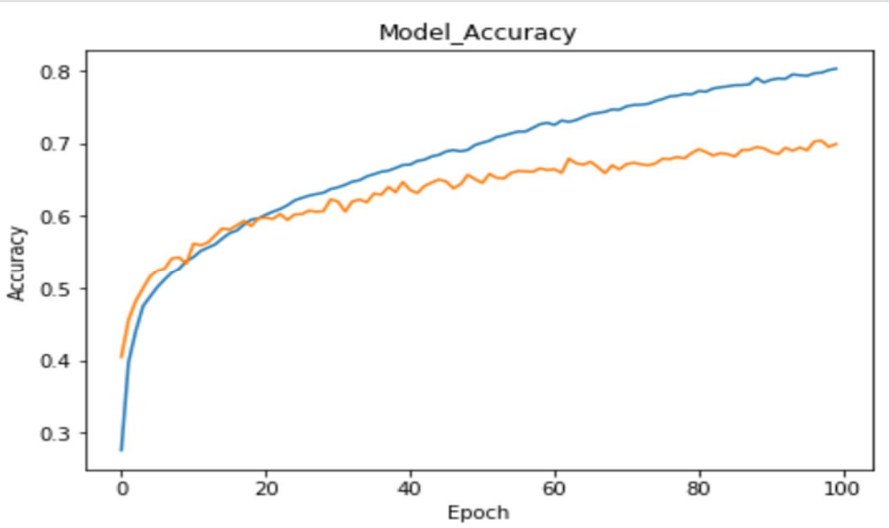


Fig: Accuracy of the Proposed model

$$\text{Accuracy} = \frac{Obtained\ Accuracy}{Expected\ Accuracy} * 100$$
$$= \frac{0.6992}{0.8027} * 100 = 87.10\%$$

This degree of accuracy suggests that the sentiment and emotion analysis algorithm can consistently categorise the majority of tweets into their corresponding sentiment categories (positive, negative, neutral), as well as emotional labels.

VII. COMPARATIVE ANALYSIS

SR.NO	ALGORITHM	FEATURES	LIMITATIONS
1	C4.5 Algorithm	• Build Models can be easily interpreted. • Easy to implement. • Can use both discrete and continuous values. • Deals with noise.	• Small variation in data can lead to different decision trees. • Does not work very well on a small training data set. • Overfitting.
2	ID3 Algorithm	• It produces the more accuracy result than the C4.5 algorithm. • Detection rate is increase and space consumption is reduced.	• Requires large searching time. • Sometimes it may generate very long rules which are very hard to prune. • Requires large amount of memory to store tree.
3	K-Nearest neighbor Algorithm	• Classes need not be linearly separable. • Zero cost of the learning process. • Sometimes it is Robust with regard to noisy training data. • Well suited for multimodal classes.	• Time to find the nearest Neighbours in a large training data set can be excessive. • It is Sensitive to noisy or irrelevant attributes. • Performance of algorithm depends on the number of dimensions used.
4	Naive Bayes Algorithm	• Simple to implement. • Great Computational efficiency and classification rate • It predicts accurate results for most of the classification and prediction problems.	• The precision of algorithm decreases if the amount of data is less. • For obtaining good results it requires a very large number of records.
5	Support vector machine Algorithm	• High accuracy. • Work well even if data is not linearly separable in the base feature space.	• Speed and size requirement both in training and testing is more. • High complexity and extensive memory requirements for classification in many cases.
6	Artificial Neural Network Algorithm	• It is easy to use, with few parameters to adjust. • A neural network learns and reprogramming is not needed. • Easy to implement. • Applicable to a wide range of problems in real life.	• Requires high processing time if neural network is large. • Difficult to know how many neurons and layers are necessary. • Learning can be slow.

VIII. CONCLUSION

To summarise, Twitter sentiment and emotion analysis using Bi-LSTM in RNN is a powerful and effective method for interpreting the sentiment and emotions represented in tweets. Data gathering, pre-processing, feature extraction, model training, evaluation, deployment, and visualisation are all stages of the proposed system. The method can provide accurate and important insights into public opinion and emotions on a wide range of topics and situations, allowing corporations, organisations, and people to make well-informed decisions based on popular opinion.

Sentiment140, DeepMoji, and LSTM-ER are three existing systems that have proved the usefulness of employing Bi-LSTM in RNN for Twitter sentiment and emotion analysis. These algorithms have demonstrated great accuracy in sentiment and emotion recognition and are widely employed in research and industry.

The use of graphs and charts to visualise information has aided researchers and analysts in gaining a better understanding of the data. However, there are certain limitations to employing Bi-LSTM in RNN to analyse Twitter sentiment and emotion. One of the most difficult aspects is dealing with sarcasm and irony, which can be difficult to detect in text. Another difficulty is dealing with noisy and unclear material, such as misspellings and informal language, which can impair analysis accuracy.

Cultural and linguistic differences can impair the accuracy of sentiment and emotion analysis, because various languages and cultures may represent sentiment and emotion in different ways. Future research might concentrate on overcoming these obstacles and enhancing the accuracy and robustness of sentiment and emotion analysis on Twitter.

Furthermore, incorporating other methodologies, such as natural language processing (NLP) and machine learning (ML), can improve the analysis's accuracy and effectiveness.

Overall, Twitter sentiment and emotion analysis utilising Bi-LSTM in RNN has a high potential for delivering useful insights into public opinion and emotions, and it has the ability to have a substantial impact on a variety of sectors such as business, politics, and social sciences. It is a fast-evolving area, and future improvements are projected to improve its capabilities and usefulness even further.

© IJRASET: All Rights are Reserved |

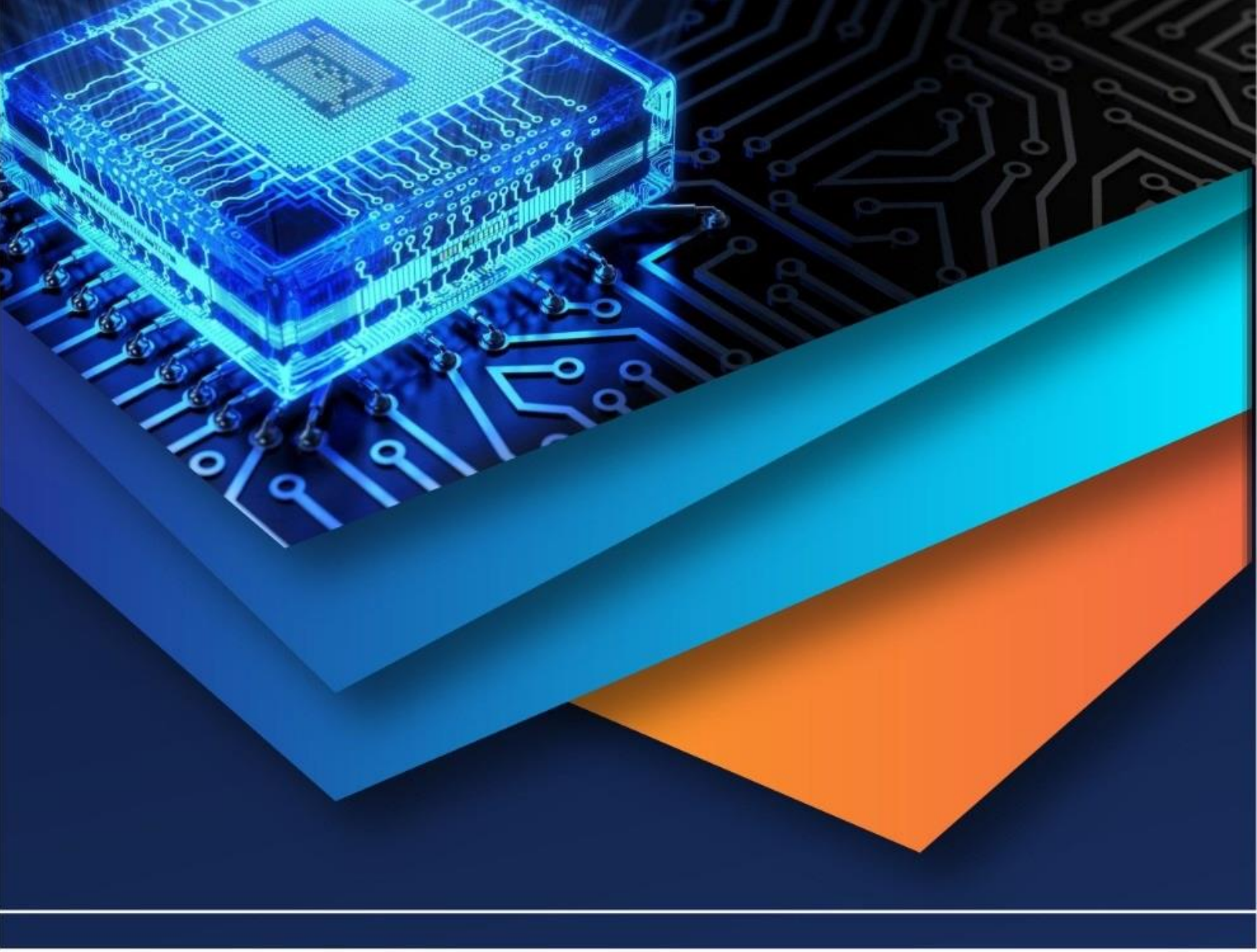
IX. FUTURE SCOPE

1) Enhance sentiment analysis capabilities by going beyond the fundamental positive, negative, and neutral sentiment categories with fine-grained sentiment analysis. Examine techniques for categorising emotions into more precise groups, such as sentiment intensity or emotion towards particular elements or things in tweets.
2) Contextual Understanding: Boost the model's comprehension and capture of tweets' contextual information. Investigate methods for integrating contextual elements into the sentiment and emotion analysis process, such as user information, conversation threads, or external knowledge graphs.
3) Analysis of sentiment and emotions exhibited towards particular features or entities mentioned in tweets using an aspect-based approach. The sentiment connected to various characteristics or aspects of goods, services, or events can be better understood using this aspect-based sentiment analysis.
4) Multilingual Analysis: The Bi-LSTM model should be modified to perform sentiment and emotion analysis across languages. Create methods to handle the issues presented by the multilingual Twitter data's different linguistic patterns, cultural quirks, and code-switching.

REFERENCES

[1] Chen, T., & Duan, Y. (2019). Twitter Sentiment Analysis using Bi-LSTM-CNN Model. Proceedings of the 2019 International Conference on Internet Computing and Big Data (ICBD), 85-89. DOI: 10.1145/3322645.3322676.
[2] Zhang, H., & Wang, L. (2020). Emotion Analysis on Twitter Data using Bi-LSTM Network. Proceedings of the 2020 International Conference on Artificial Intelligence and Computer Science (ICAICS), 279-283. DOI: 10.1145/3429486.3429493.
[3] Felbo, B., et al. (2017). Using Millions of Emoji Occurrences to Learn Any-domain Representations for Detecting Sentiment, Emotion, and Sarcasm. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1615-1625. DOI: 10.18653/v1/D17-1169.
[4] Jindal, A., & Liu, B. (2008). Opinion Spam and Analysis. Proceedings of the International Conference on Web Search and Data Mining (WSDM), 219-230. DOI: 10.1145/1341531.1341557.
[5] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751. DOI: 10.3115/v1/D14-1181.
[6] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.
[7] Zhang, X., & LeCun, Y. (2015). Text Understanding from Scratch. Proceedings of the 2015 International Conference on Machine Learning (ICML), 1242-1250.
[8] Alsmadi, I., et al. (2018). Sentiment Analysis of Twitter Data: A Comprehensive Review. International Journal of Computational Linguistics Research, 9(2), 69-82.
[9] Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC), 1320-1326.
[10] Go, A., et al. (2009). Twitter Sentiment Classification using Distant Supervision. Proceedings of the 23rd International Conference on Computational Linguistics (COLING), 2, 2009, 401-408.

© IJRASET: All Rights are Reserved |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)