IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

TEXT SENTIMENT ANALYSIS USING LSTM

Riya*1, Gaurav*2, Manish Kumar Gupta*3, Bhavya Tyagi*4, Jay Shankar Prasad*5

*1,2,3,4Department of Computer Science Engineering, GNIOT, Greater Noida, India

DOI : https://www.doi.org/10.56726/IRJMETS38959

ABSTRACT

Sentiment analysis is the study of extracting opinion from expressed language. It comprehensively examines data stored on the Internet to identify and categorise the views expressed in a text. Frequently, people express their opinions or views on various events in society or the world by writing textual reviews. The current work focuses on evaluating the reviews in terms reviewer's opinion and categorising them as a positive, negative and neutral.

To find the sentiment from the reviews, well known machine learning techniques Long Short Term Memory (LSTM) and recurrent neural Network (RNN) are used and their performance is also evaluated. For experimentation, reviews data from Kaggle is used and the LSTM model achieves an accuracy of around 78%.

I. INTRODUCTION

Analyzing sentiments of people has become increasingly important in today's world. Analyzing sentiments of people plays an important role in understanding customer opinions and feedbacks to improve products, services, and customer experience. Brand Reputation Management by monitoring and analyzing online conversations and social media posts which can help the companies to timely address all negative reviews about them, Marketing and Advertisement, Crisis Management, and even in Political and Social Analysis which can be helpful in reducing negative effects of Social unrest and provide valuable insights for policymakers and experts. But, these models are made to work for a specific domain, for example, Market Specific, Movie or other sentiment Analysis models, so there is a need for Generalized Sentiment Analysis models which is trained on data from different domains. Due to growing use of social media,Business and data analysts can use this data to gauge brand reputation and understand customers requirement and feedbacks, this is why analysing sentiments is also referred as Opinion Mining, as in today's world a single negative post or tweet about a business can ruin its reputation, therefore it is very important to monitor people's opinion, not only that but it can be helpful in gaining valuable insights by analyzing customer's preferences and feedbacks, product analysis and can provide innovative ideas, and play a great role in future growth prediction for any organization. It can be helpful in monitoring social unrest, which can reduce riots to a great extent. Analyzing sentiments mainly focuses on identifying or extracting subjective information, to get an idea of overall response of people, i.e., it is the overall summary of reviews or responses of people. Our main objective is to build a system that is able to extract the subjective information or author's attitude towards a particular topic, product, etc., and summarize it as a positive, negative, or neutral response. The system will also summarize the percentage of these responses for better analysis.

II. LITERATURE REVIEW

The sentiment analysis has a long history, in 1970, psychologist Paul Ekman and his colleagues Wallace Friesen published their study on the Universal Expression of emotions [1]. It was based on the hypothesis stating that there are six universally recognized emotions: surprise, anger, fear, disgust, sadness, and happiness. For testing the hypothesis, a total of around 30 photographs displaying facial expressions of these six emotions were shown to people living in Brazil, The United States, Argentina, Chile and Japan to predict the emotions or sentiments associated with the expression. The results supported the hypothesis proving these six emotions are universally recognized. The study proved that certain facial expressions are universally recognized. Their study was further turn out to be useful in different areas such as emotion recognition, mental health etc. Maheshwari Selvaraj research focused on developing efficient and accurate models to recognize emotions from speech signals. This research involves finding acoustic features which can be used in extracting sentiments of the speaker [2]. This research utilizes techniques like feature extraction, signal processing for analyses of speech signals and recognize emotions. This research also address key challenges associated with speech based sentiment analysis, which are missing data due to signal loss, inter and intra speaker variability. In [3], authors used a multi-modal approach by combining facial expressions and speech signals, for emotion detection. The feature extraction from both the speech signals and facial expression of the participants while they were watching emotional video clips were used. Features used to train Emotion recognition models. The results shown in [3] is Bimodal systems results and having higher accuracy than Unimodal systems. In [4], a multi-modal sentiment analysis by using Electroencephalogram (EEG) signals with audio and visual modalities discussed. The feature extraction was performed on EEG, speech, and facial signals, that were recorded while participants watched an emotional video clip. The extracted features used to train Emotion recognition models. The results produced was encouraging.

In [5, 6], the authors compared the performances of Support Vector Machines (SVMs), Random Forest, Naïve Bayes, and Decision Trees machine learning algorithms for the analysis of sentiments. The impact of feature selection techniques term-frequencyinverse document frequency (TF-IDF) were also evaluated in [5]. Results of the study showed in [5, 6] SVM algorithm performed best on the data set consisting of hotel reviews and product reviews. This research proved the importance of selecting right machine learning algorithm, suitable pre-processing and feature selection technique. While performing the sentiment analysis, the missing value is prime concern which arises many times due to unavailability of data, to handle this authors in [7], Used an Iterative data augmentation to generate synthetic data for filling the gaps in the place of missing values Multiple Machine learning models trained on specific modality are used to predict missing data. Repetition of this process occur iteratively until satisfactory level of performance is gained. This approach is mainly focuses on addressing missing modality problem in sentiment analysis.In [8], aspect-based analysis of sentiment using new target representation and dependency attention discussed. The dependency attention mechanism solved using graphical neural networks to model dependency between words in text. These graph neural networks contain information from both word and target embedding. The study of [8] showed the importance of incorporation of target representation and dependency attention mechanism in aspect based analysis of sentiments and to achieve more accurate and robust models.

III. METHODOLOGY

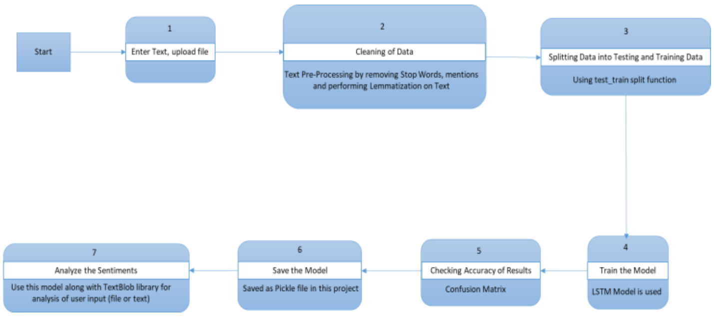

Figure 1: Sentiment analysis model

Figure 1 is the sentiment analysis model used it takes as input a Multi-class Sentiment datasets containing product reviews taken from Amazon, Tweets, etc., to train model. In the dataset, the training data is labelled as positive, negative and neutral. The first step is to clean the data and the Data Cleaning process involves pre-processing of data, i.e., removal of stop words like as, and, or, as these words do not contribute much in analysis and removing them reducing data size. Further cleaning involves removing all strings or text having pattern like, '@any_string'. Lastly, lemmatization of words is performed which converts the words into meaningful base form or lemma. The second step is Categorical Encoding, in this step, the numerical data is converted into categorical one, as machines can only understand numbers. Here, mapping of each sentiment like, for category positive, it is mapped to number 0, for negative it is mapped to 1, and for neutral it is mapped to 2. This step ensures that categorical data is treated as distinct values, rather than arbitrary numerical values which might lead to unwanted relationships or bias in data. Here, we are only converting the labels associated with each sentence in the dataset into numerical data. The third step is Preparing Data for Model, in this step all textual data, i.e., all sentences in the dataset are converted to a sequence of integers. All sequences have same length, which is equal to the length of longest sequence in the training dataset. The fourth step is Neural Network Architecture, here the LSTM (Long Short Term Memory) is used for the processing of the data. LSTM layer contains 32 memory cells or hidden units. Output of Embedding layer is taken as input by each layer. Embedding layer just learns the vector or integer representation for each word. After this mapping of vector of arbitrary real valued numbers to probability distribution. The softmax activation function is used. Here, the model learns to adjust its weights so that it can accurately predict the output of input data, here since the labels for input training data are already there so, the model just adjust weights to give almost same result as that of the already depicted results. After training phase, model weights are fixed (they are fixed or set as the one which gives highest accuracy for input data). The model then uses these learned weights to process test data and provide results, which are to predict whether sentence is positive, negative or neutral.

IV. RESULTS AND ANALYSIS

The study initiated with user inputs, which can be a file, text, or even can search for a specific tweet and time range to which the tweet must belong to. For each type of input mentioned above there is separate page, user can choose option available in the menu according to type of his/her input data, as shown in Fig. 2. For file option, user can upload the file or drag the file, the results will be each sentence in the file is shown along with its label, which is one of positive, negative or neutral, a Pie Chart depicting the percentage of positive, negative, and neutral sentences in the file will also be shown, Fig 3. is displaying results of an input file. For textual data, the user can write or copy paste the input data and will get results as label positive, negative and neutral, as shown in Fig. 4. For analyzing tweets, user just need to enter the word to search and then the timestamp within which the tweets must range, the results are same as that in file option

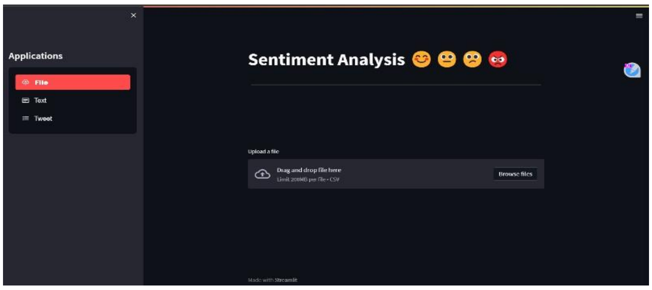

Fig. 2. User Interface

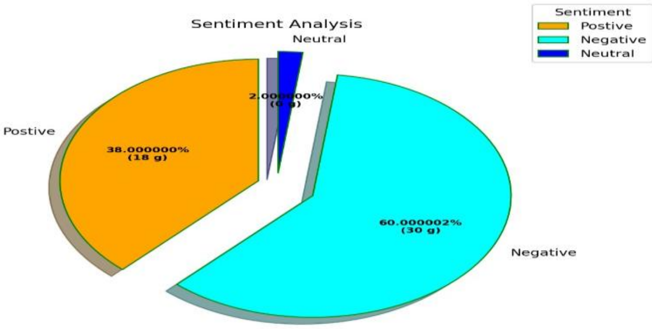

Fig. 3. Output Interface for File page

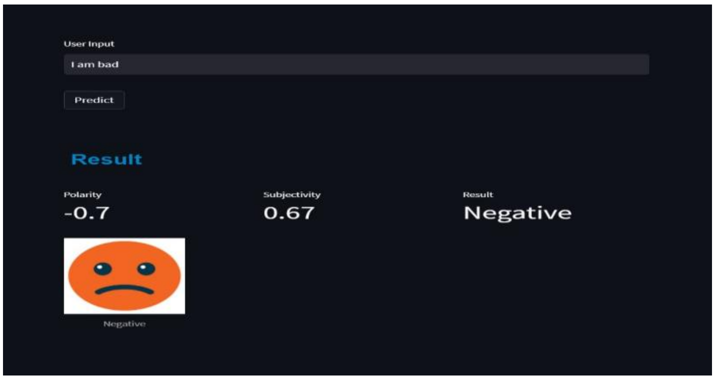

Fig. 4. Output Interface for Text Page

V. FUTURE SCOPE

The sentiment analysis can be used for various application including the monitoring of social media apps at real time. In future, analysing the sentiment along with the suggestions may solve many problems, for example, if most of the reviews for a mobile company are about its lower battery backup, then it can analyse the things and provide solution to company by suggesting how to solve this issue with resources in hand. For understanding and predicting political opinions and polls, keeping a check on social sites, preventing public review crises, increasing morale and boosting productivity by listening to the employees, predict and analyses market trends. Analysing large amounts of employee feedback data to determine employee satisfaction levels can also be a future work.

VI. CONCLUSION

In this paper, labelled data to design a model for analysis of text sentiments used and we extracted the subjective information from written language which can further be used in business or other analysis to understand people's opinions, sentiments, attitudes, and emotions. It is able to predict the sentiments of the text and label them into multi-class sentiments, which are 'Positive', 'Negative', and 'Neutral'.

VII. REFERENCES

[1] Paul Ekman and others: Universal facial expressions of emotion, Culture and Personality: Contemporary Readings/Chicago, pp. 151–158, 1974.

[2] S. Maheshwari , R. Bhuvana, S. Padmaja, Human Speech Emotion Recognition, / International Journal of Engineering and Technology (IJET), Vol 8 No 1, pp.311-323, p-ISSN : 2319-8613

[3] L. C. De Silva and Pei Chi Ng, "Bimodal emotion recognition," Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 2000, pp. 332-335, doi: 10.1109/AFGR.2000.840655.

[4] L. S. Chen, T. S. Huang, T. Miyasato and R. Nakatsu, "Multimodal human emotion/expression recognition," Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 366-371, doi:

[5] 10.1109/AFGR.1998.670976.

[6] A. Alshamsi, R. Bayari, S. Salloum "Sentiment

[7] Analysis in English Texts", Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 6, pp. 1683-1689 (2020).

[8] X. Fang, J. Zhan, Sentiment analysis using product review data. Journal of Big Data 2, 5 (2015). https://doi.org/10.1186/s40537-015-0015-2

[9] Wang, Ning. "M2R2: Missing-Modality Robust emotion Recognition framework with iterative data augmentation." ArXiv abs/2205.02524 (2022): n

[10] T. Yang, Q. Yin, L. Yang , O. Wu, Aspect-Based Sentiment Analysis with New Target Representation and Dependency Attention,IEEE Transactions on

[11] Affective Computing, vol. 13, no. 2, pp. 640-650, 1

[12] April-June 2022, doi: 10.1109/TAFFC.2019.2945028.