ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

BIG DATA SENTIMENT ANALYSIS FOR PREDICTING FOUR CATEGORIES OF EMOTIONS USING BILSTM ALGORITHM

1. Rohan Steven B
Department of Electronics and Communication Engineering
Meenakshi Sundararajan Engineering College
rohansteven0904@gmail.com
https://orcid.org/0009-0002-8059-041X

2. Soorya Narayanan V
Department of Artificial Intelligence and Data science
Meenakshi Sundararajan Engineering College
soorya220104@gmail.com
0009-0009-0605-9509

3. Sandhya C
Department of Electronics and Communication Engineering
Meenakshi Sundararajan Engineering College
sandhyachengaiyan@msec.edu.in
0000-0001-6236-1491

4. Ramani S
Department of Artificial Intelligence and Data science
Meenakshi Sundararajan Engineering College
sramani@msec.edu.in
https://orcid.org/0000-0003-0240-137X

5. Sivakumar J
Department of Electronics and Communication Engineering
St.Josephs's College of Engineering
sivakumarj@stjosephs.ac.in
https://orcid.org/0000-0001-8964-6032

ABSTRACT:
Sentimental analysis is a way of understanding human text expressions based on the lexical content. It is an important in natural language processing (NLP) and machine learning that allows researchers and marketers to analyse public sentiment and understand customer sentiment from textual data. This work uses Bidirectional Long Short-Term Memory (BiLSTM) neural networks along with Adam optimizer to classify records of big data (31017 samples) into three types of sentiments and also the unknown such are non-expressive texts in the social media. With an entire dataset from tweets, the preprocessing steps includes text cleaning, tokenization, and lemmatization to improve the model's performance. With the goal of achieving high category prediction, the architecture includes embedding layers, BiLSTM layers for capturing long-term memory of the given dataset, and a dropout for regularization. Metrics for evaluating the models

include, recall, accuracy, precision, F1 Score and receiver operating characteristic (ROC) curve. These metrics collectively establish how reliable sentimental prediction is at a given point in the research process. The validity of the tool lies in its capability to provide emotion categorization into positive, neutral and negative with superior accuracy and to meet the stressful conditions of emotion evaluation from the whole samples. By incorporating these advanced techniques, our system provides an actual-time sentiment analysis, offering advanced functionality for social media management like Twitter, and each other area studying emotional intelligence. The model also consists of an extensive evaluation for overall performance and model enhancement, with a view to ensuring transparency and reliability in sentimental classification results.

Keywords—NLP, Sentiment Analysis, Big data, BiLSTM

INTRODUCTION:

Sentiment analysis is one of the important type in analysing social media, consumer understanding and research interpretation [9]. Provides valuable insights into prevailing sentiment towards products, services, and topics [11]. It facilitates informed decision-making based on the lexical lines and strategic planning for the improvement in their respective fields [10]. In this work, sentiment analysis is applied to a dataset of tweets, which is a representative sample of public opinion expressed in the general language with symbols and non-alpha characters [13, 17]. The main objective of the task was to classify these tweets into three emotion categories namely positive, neutral, negative and non-emotional texts. Textual correction is followed by lemmatization, where words are reduced to their original form in order to improve the accuracy of the text and reduce distortion [16]. They eliminate stop words, which are common but less informative words, to focus on important words that help convey emotion. To verify the accuracy with the input data, the clean text is subsequently tokenized into a numerical sequence and padded [15]. Preprocessed data is analysed using the Bidirectional Long Short-Term Memory (BiLSTM) network. The bidirectional LSTM layers in the model architecture establish connections between the sequences after an embedding layer for transforming the text into dense vectors [4]. Dropout layers stop overfitting, whereas thick layers with ReLU and softmax activations display three sensitivity levels and the unknown [3].

The analysis of sentiments using the BiLTSM model for the dataset which contains three sentiments and unknown which is well understood by the model and classifies is into positive, neutral, negative and unknown (non-emotional texts) [5]. All the classification is found using tweets which a count of 31017 samples which comes under big data analysis.

This data samples are given to the model to test and train and it is made up of tweets that have the users emotional states [1]. The dataset is classified into three main sentiments. Positive emotions are the thoughts and sentiments that are conveyed through the tweets, while negative emotions are melancholy or adverse responses. Neutral emotions are those seen in tweets that are not expressive like positive and negative feelings. Finally unknown which doesn’t express anything. These differences between the sentiments provide the feelings of the public opinion and widen a more thorough analysis of discussions on certain topics thus to extract the meaningful expression from the textual data and giving a beneficial tool for understanding public opinion and making decisions [1].

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

The amazing improvement in text emotion evaluation was validated by the creation of Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks [3]. The LSTM community to address the issue of vanishing gradients in well-known RNNs, making it suitable for coping with lengthy sequences [7]. The use of LSTM networks in sensitivity analysis, allowed better processing of context and sequential dependencies and stepped forward overall performance in informational emotional expressions in complicated sentences [2].
Recently, Pedro M. Sosa (2017) proposed two neural network models to improve the sentimental analysis on Twitter data and finally concluded with the improvement around 2.7%-8.5% better than other regular models [19]. Akana Chandra Mouli Venkata Srinivas (2021) performed sentimental analysis using LSTM-CNN on the social media conversation with the accuracy of 87% for predicting positive and negative only [20].
From the above citations it is clear that research or papers are mostly for predicting positive and negative only or performing big data sentimental analysis on the same, even some papers have worked on three sentiments with minimal data samples [17]. In this work, big data sentimental analysis is performed to predict four category such as positive, neutral, negative and unknown sentiments.

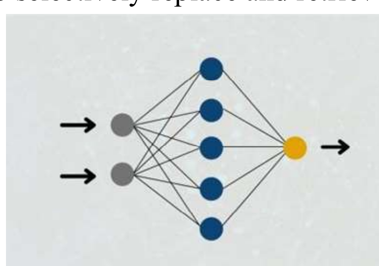
BACKGROUND:

Recurrent Neural Networks (RNNs) Background:

RNN models are designed to deal with sequential records by keeping a hidden state that captures temporal dataset [8]. Despite their capability, RNNs conflict with analyzing long-time period dependencies due to vanishing gradients during backpropagation through time (BPTT). This obstacle makes them wrong for obligations requiring reminiscence over lengthy sequences.

LSTM Architecture:

LSTMs deal with this problem by means of introducing a memory cell that can maintain its state even for lengthy sequences of data. The architecture controls records using three gates (forget, input, and output), allowing it to selectively replace and retrieve information [6].

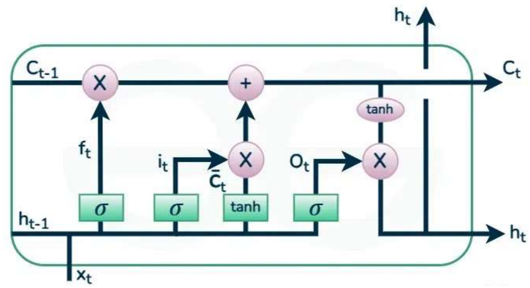


1.Forget Gate: This gate is used to select the previous memory elements to retain.

2.Input Gate: This gate is used to determine which new information should update the memory.
3.Output Gate: Controls the output from the memory cell to the following layer or time step.
Each gates have its own set of weights and biases, which is observed during training. The aggregate of those gates offers LSTM its capacity to deal with long-term memory successfully.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

Derivation of LSTM Equations



At each step t, the LSTM unit performs a set of operations to update its cell state and hidden state.

1. The forget gate controls how much of the previous cell state is retained:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

where Wf and bf are the weight matrix and bias for the forget gate, ht−1 is the hidden state from the preceding time step, and xt is the current input.

2. The input gate controls the influence of the current input on the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

The new records to be introduced to the cell state is:

$$\hat{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

3. The cell state is updated using the forget gate and the input gate:

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{C}_t$$

Where:

- $\odot$ element-wise multiplication
- tanh activation function

This equation enables the cell state to keep long-term records at the same time as moreover updating with new data.

4. Finally, the output gate generates the hidden country for the following time step:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

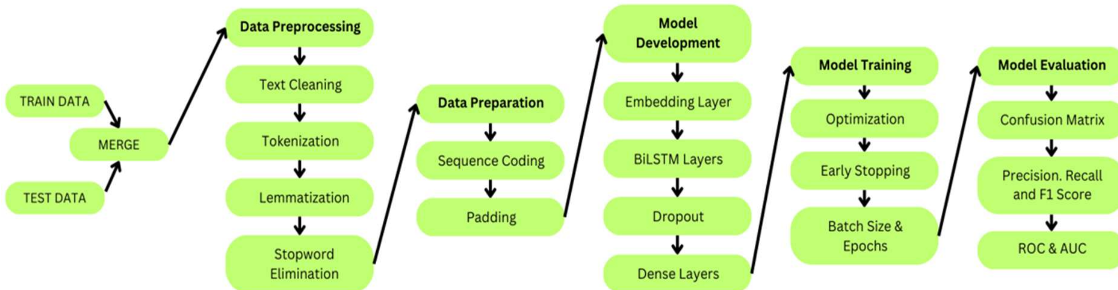
The new hidden kingdom is computed as:

$$h_t = o_t * \tanh(C_t)$$

The gate mechanism in LSTM networks allows the model to control the sudden variation of facts, making it suitable for predictions regarding long-term memory state [12]. By leveraging those developments, LSTMs have emerge as a foundational device in modern day to day tool for analyzing applications [6].

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

METHODOLOGY:



The method for the analysing emotional text assessment project employs an intensive and methodical technique that consists of preprocessing, sampling, training, test and evaluation. This approach guarantees that the accuracy of evaluation version is robust, unique, and sensible for human textual content.

1.Data Preprocessing:

This work is built on twitter dataset that contains the original text without skimming the records. The training and testing dataset is constituted of two fantastic CSV files, "train_centi.csv" and "test_centi.csv" which is joined together to make a count of 31017. Every detail within the collection has a tweet at the aspect of the sentiment label related to it. Preparing the information ensures that it is smooth and suitable for the model to learn. It is a vital phase in the process.

```
Sample Counts:
        Sample Type  Count
0     Total Samples  32296
1  Training Samples  29066
2   Testing Samples   3230
```

The method involves:

Text Cleaning: To remove useless text and unwanted characters, tweets are first wiped clean. This includes doing away with HTML tags that may seem due to the delivery of the tweets, in addition to non-alphanumeric and alphanumeric characters, which do no longer contribute to the meaning [6]. To make the textual content more uniform, the extra white space is also been removed. Based on their substantial content, special characters are eliminated from tweets so that it will save you tampering with the tokenization procedure [13].

Tokenization: The extracted textual content is splited up into separate phrases, or tokens. Tokenization is critical for modelling textual input right into a format that the deep getting to know model can recognize and look at [9]. By giving each word a unique integer, a chain of integers is formed, and these integers collectively form the distinctive text [15].

Lemmatization and Stopword Elimination: Stopwords like "and" and "the" are eliminated since they are common expressions that no longer have a clear meaning and will generate noise. Phrases may be lemmatized to come to be their bases or roots (e.g., "running" becomes "run"). This method reduces the number of particular tokens, standardizing terminology while additionally enhancing the model's overall performance [16].

2.Data preparation:

Sequence Coding: Tokenizer magnification transforms separated textual content into an integer series. Each particular word in the vocabulary is assigned an integer, and each tweet is

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

represented as a chain of these integers. This degree of computation is crucial for transmitting transcriptional statistics to the neurons [3].

Padding: This line is used to make sure that all the sequences are of the same length which involves adding zeros to shorter sequences to make all the dataset to have uniform length. The model requires a constant input size to process the data correctly [11].

3. Model Development:

Creating a Bidirectional Long Short-Term Memory (BiLSTM), a type of Recurrent Neural Network (RNN) which is mostly used for sequential prediction:

Embedding Layer: Integer sequences are converted into dense vectors of constant length. During training, these layers learn word embeddings for capturing the relationship between words [18]. The model can understand the text and it's meaning of each word in relation to others by mapping words to dense vectors.

Bidirectional LSTM layers: To capture dependencies in both forward and backword directions within the text, two Bidirectional LSTM layers are used. This bidirectional LSTM layers allows the model to learn from both past and future context by providing more understanding of the text sequences [4]. The first layer is used for processing the sequence while maintaining the order of words and second layer is used for refining this understanding.

Dropout: To eliminate overfitting dropout layer is included. During training, a fraction of the input devices is randomly set to 0, which allows the model to generate better results by reducing its reliance on the specific features [16]. This regularization technique is used in unseen data to improve the model performance.

Dense Layers: This network has many dense layers along with ReLU activation functions. This layer is used to learn complex patterns in the data by adding non linearity to the model [8]. The final layer uses softmax activation to produce a probability distribution over the sentiment classes (positive, neutral, negative) and allows the model to make prediction about the sentimental of each tweet.

4.model Training:

The version requires optimizing its parameters to reduce the loss function:

Optimization: This is done using Adam optimizer which is used to adjust the model's weights and reduce the categoric cross-entropy loss function. Adam optimizer is best for this task due to its adaptive learning rate and ability to handle sparse gradients effectively.

Early Stopping: Early Stopping is implemented to monitor the validation loss during training. This loss does not improve for a specific number of epochs thus the training is halted to prevent overfitting. The best performing model weights observed during training is retrained.

Batch Size and Epochs: Batch size of 8 is used for training, meaning that 8 samples are processed before updating the model's weights. Training is conducted over 4 epochs which means four complete pass through the entire dataset. The use of both batch size and epochs will balance the training efficiency and model performance.

5.Model evaluation:

The evaluation of model's performance in crucial to understanding its effectiveness:

Accuracy: The overall accuracy of the model is calculated by finding the proportion of correctly classifying the sentiments among all the predictions. Accuracy of the model tells us who well

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

the model is performing the sentimental analysis.

Confusion Matrix: A confusion matrix is generated to visualize the model performance whether true positives, true negative, false positive and false negative. This matrix helps to understand how well the model distinguish between positive, neutral, negative and unknown.

Precision, Recall, and F1 Score: The metrics are computed from the confusion matrix to evaluate the model performance for each sentimental class. Precision measures the true positive among all the positive, recall is used to identify the model's ability to identify all the relevant instances in the dataset and F1 score will give a single matric by combining precision and recall. All these matrices will give the models performance beyond overall accuracy.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ROC Curve and AUC: Receiver Operating Characteristic (ROC) curve and the Area Under the Curve are also used for evaluating the model's ability to distinguish between sentiment classes. ROC is used to plot the curve between true positive and false positive rate, while AUC is used to plot the overall performance. These matrices are used for understand the model's performance in multi class classification.

RESULT AND DISCUSION:

In this work, we examine the overall overall performance of a sentiment evaluation version based totally on Long Short-Term Memory (LSTM) structure [14]. The metrics evaluated consist of Precision, Recall, F1 Score, Accuracy, and Sensitivity. These metrics provide a complete understanding of the way well the model performs in classifying sentiment lessons primarily based totally on a given dataset.

Result Table:

	Metric	Value
0	Accuracy	0.736842
1	Precision	0.735937
2	Recall	0.736842
3	F1 Score	0.736202

Precision:

Precision refers to the proportion of extraordinary predictions which have been in truth accurate. In the context of this sentiment evaluation assignment, it approach the percentage of instances that the version predicted as a positive sentiment (e.g., positive or negative) and have been truly

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

of that sentiment. A precision score of 0.7359 indicates that 73.59% of the version's positive sentiment predictions have been accurate.

A precision rating of 0.7359 indicates a reasonably dependable model that makes exceptionally few fake-effective predictions. The slightly decrease precision (as compared to a genuinely perfect score of 1.0) could imply that the model is rarely classifying neural or negative sentiments as positive, ensuing in false positives.

Recall:
Recall (or sensitivity) measures the proportion of actual outstanding times that have been efficaciously identified by using the model. A recall score of 0.7368 means that 73.68% of the positive sentiments inside the given test dataset were successfully diagnosed via the model.

This rating shows that the version has an awesome capability to seize most of the positive sentiment instances within the dataset. However, a recall score slightly less than 1.0 indicates that a few exquisite sentiments could have been neglected, that means there are some fake negatives.

F1 Score:
The F1 score is the harmonic mean of precision and recall, offering a stability some of the two metrics. It provides a single metric that summarizes the model's usual performance when precision and recall are equally important. In this work, the F1 score is 0.7362, because of this that the stability between precision and recall is steady, with each metrics contributing similarly to regular version overall performance.

The F1 score of 0.7362 indicates that the version keeps a constant typical performance in terms of balancing precision and recall. A good F1 score close to the values of precision and recall implies that there aren't super discrepancies among how nicely the version handles false positives and fake negatives.

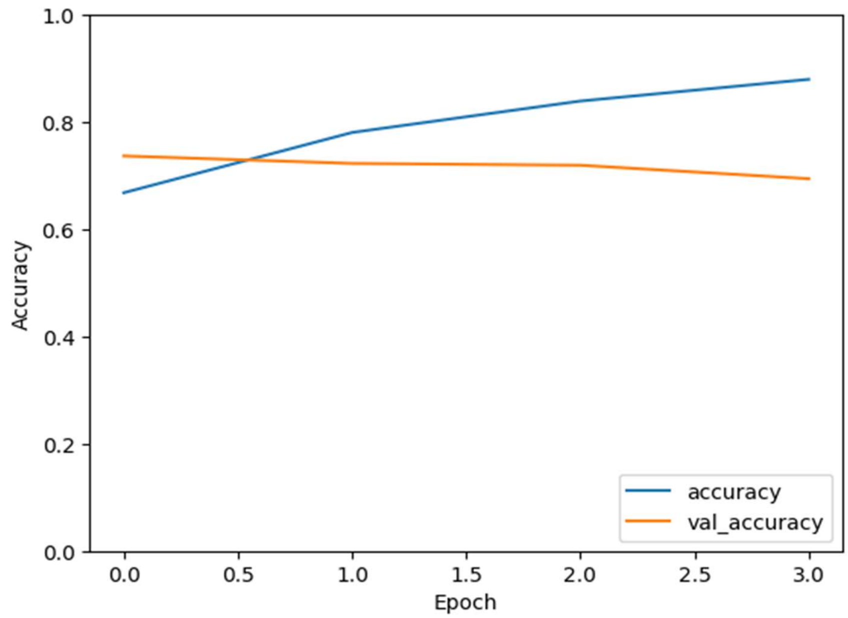
Sensitivity:
Sensitivity (referred to as recall, particularly in binary classification) is the percentage of real positives which may be correctly categorised. In this work, sensitivity is equal to 0.7368, similar to recall.

Sensitivity is especially crucial in this work because we want the model to be extraordinary in detecting positive sentiment effectively. The sensitivity rating shows that 73.68% of the positive instances had been correctly predicted.

Accuracy:
Accuracy measures the share of successfully categorised instances out of the whole quantity of instances. In this model, an accuracy of 0.7368 approach that 73.68% of the total times were efficiently categorised across all sentiment classes.

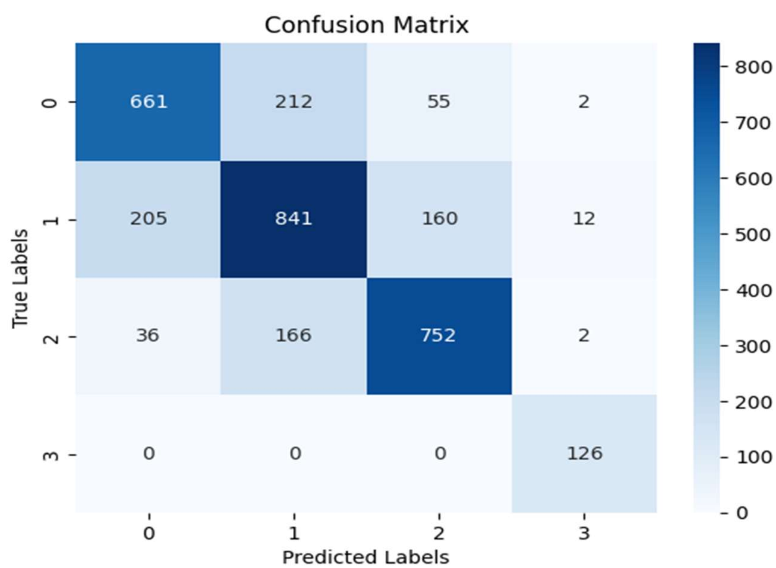
ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

An accuracy of 73.68% suggests that the model plays properly throughout the whole dataset. However, accuracy alone can't produce the exact score and sometimes can be misleading, especially in imbalanced datasets. In this work, given that precision and recall are every close to the accuracy, we are able to infer that the model within reason balanced in its class of differentiating sentiment commands.



Confusion Matrix:

The confusion matrix gives an outline of the performance that the model completed in terms of correctly and incorrectly classifying sentiments. It helps in identifying where the model struggled, whether with false positives or false negatives.



ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

From this matrix, we've got a look at the following styles:
True Positives (TP): The model correctly labelled a categorized as high number of positive sentiments.
False Positives (FP): The model rarely predicted positive sentiment as neutral or negative. False Negatives (FN): A tremendous number of actual positives were incorrectly classified as neutral or negative.

Class breakdown:
Positive class (class 0): 661 tweets were correctly classified as positive, but it also classified 212 tweets as neutral and 55 tweets as negative. There are a few cases where a good tweet was predicted without knowing it.

Neutral class (class 1): the prediction of this class is relatively strong, by correctly classifying 841 tweets as neutral however, 160 tweets are classified as negative and 205 as positive.

Negative class (class 2): 752 tweets are correctly classified as negative, but 166 tweets of positive and 36 tweets of neutral is misclassified as negative.

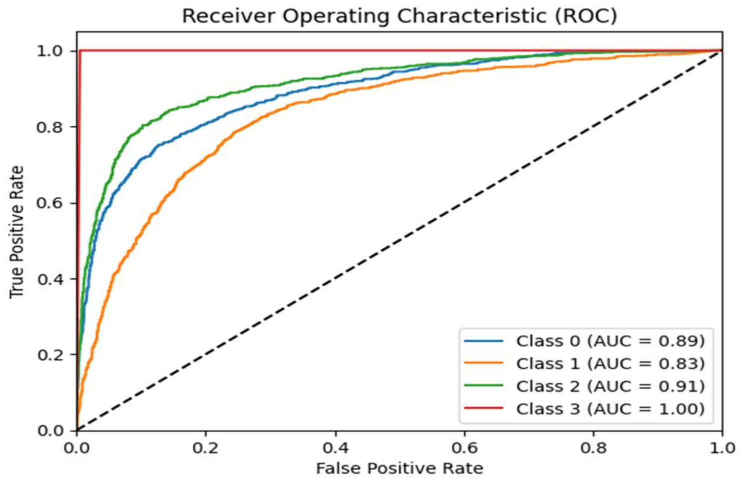
Unknown class (class 3): the model classified the unknown tweets correctly without any misclassification.

Receiver Operating Characteristic (ROC) Curve:
The ROC curve for every sentiment tells the trade-off among true positive rate (sensitivity) and false positive rate. The Area Under the Curve (AUC) for every sentiment gives how well the model differentiates among classes:

AUC for Positive Class: 0.89
AUC for Negative Class: 0.91
AUC for Neutral Class: 0.83

Discussion:
An AUC above 0.7 throughout all classes shows that the model is incredible however not perfect ability to differentiate among different sentiment categories. The ROC curve analysis in addition confirms that the model plays similarly across all sentiment instructions.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)



CONCLUSION:

The model is successfully trained using Bidirectional Long Short-Term memory (BiLSTM) network showcasing the model's efficiency in sentimental analysis. The model is tested and trained on the tweets dataset in the ratio on 1:9 over 31017 samples and produced an accuracy of 73.16% by preprocessing and training tweets dataset. This rate of accuracy shows how well the trained model classify the sentiments throughout special contexts. The confusion matrix found that the model has trouble in identifying the sentiments positive, neutral and negative accurately but finds the unknown text with 100%. The key metrics Precision (73.55%), Recall (73.16%) and F1 score (73.02%) reflects the model's balanced performance across the sentimental classes.

The result shown using BiLSTM algorithm classifies the positive, neutral and negative with the accuracy of 73.16% and along with this we have also found the unknown texts that doesn't contain any emotion. In this work the model is trained to classify totally four category where are positive, neutral, negative and unknown texts from the dataset of 31017 samples.

In future, enhancing the model's performance to differentiate between closely related sentiments to potentially improve the overall accuracy. Further improvement in fine tuning hyperparameters or using an advanced algorithm architectures to better understand the sentiments in the text and finally exploring different types of methodology to improve the model's accuracy and showcasing the comparison of other models on the domain sentimental analysis to have better understanding on human emotions through texts.

REFERENCE:

1. Sharma, T., Diwakar, M., Singh, P., Lamba, S., Kumar, P., & Joshi, K. (2021). Emotion Analysis for predicting the emotion labels using Machine Learning approaches. *2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. https://doi.org/10.1109/upcon52273.2021.9667562

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

2. Tang, D., Qin, B., & Liu, T. (2015). Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. *Conference on Empirical Methods in Natural Language Processing*.

3. Aziz, M. A., Islam, M. M., Munshi, M. S. P., Samshi, S., & Khan, M. M. (2022). Sentimental Analysis on reviews of Protein Supplement using RNN-LSTM. *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*. https://doi.org/10.1109/iccmc53470.2022.9753901

4. Venkatesan, D., Kannan, S. K., Arif, M., Atif, M., & Ganeshan, A. (2022). Sentimental analysis of Industry 4.0 perspectives using a Graph-Based Bi-LSTM CNN model. *Mobile Information Systems*, *2022*, 1–14. https://doi.org/10.1155/2022/5430569

5. Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, *14*, 100157. https://doi.org/10.1016/j.array.2022.100157

6. Pavithra, E., Geetha, A., & Shri, B. S. (2024). Correlational and sentimental analysis using LSTM to forecast stock price. In *CRC Press eBooks* (pp. 365–376). https://doi.org/10.1201/9781003430452-54

7. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

8. Shilpa, P. C., Shereen, R., Jacob, S., & Vinod, P. (2021). Sentiment analysis using deep learning. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. https://doi.org/10.1109/icicv50876.2021.9388382

9. Rahman, S. E., & Ullah, S. (2020). Email Spam Detection using Bidirectional Long Short Term Memory with Convolutional Neural Network. *2020 IEEE Region 10 Symposium (TENSYMP)*. https://doi.org/10.1109/tensymp50017.2020.9230769

10. Shaikh, A., Panuganti, A., Husain, M., & Singh, P. (2021). Stock market prediction using machine learning. In *Advances in intelligent systems and computing* (pp. 497–508). https://doi.org/10.1007/978-981-15-8443-5_42

11. Saha, U., Mahmud, M. S., Shimu, S., Eva, S. N., Asif, I. A., & Khushbu, S. A. (2022). The corporeality of infotainment on fans feedback towards sports comment employing convolutional Long-Short term neural network. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. https://doi.org/10.1109/icoei53556.2022.9777216

12. Xia, C., Zhao, D., Wang, J., Liu, J., & Ma, J. (2018). ICSH 2018: LSTM based Sentiment Analysis for Patient Experience Narratives in E-survey Tools. In *Lecture notes in computer science* (pp. 231–239). https://doi.org/10.1007/978-3-030-03649-2_23

13. Gutub, A., Shambour, M. K., & Abu-Hashem, M. A. (2023). Coronavirus impact on human feelings during 2021 Hajj season via deep learning critical Twitter analysis. *Journal of Engineering Research*, *11*(1), 100001. https://doi.org/10.1016/j.jer.2023.100001

14. Biradar, S. H., Gorabal, J., & Gupta, G. (2022). Machine learning tool for exploring sentiment analysis on twitter data. *Materials Today Proceedings*, *56*, 1927–1934. https://doi.org/10.1016/j.matpr.2021.11.199

15. Sharma, A., & Ghose, U. (2020). Sentimental Analysis of Twitter Data with respect to General Elections in India. *Procedia Computer Science*, *173*, 325–334. https://doi.org/10.1016/j.procs.2020.06.038

16. Das, D. D., Sharma, S., Natani, S., Khare, N., & Singh, B. (2017). Sentimental Analysis for Airline Twitter data. *IOP Conference Series Materials Science and Engineering*, *263*, 042067. https://doi.org/10.1088/1757-899x/263/4/042067

17. Jain, D. K., Boyapati, P., Venkatesh, J., & Prakash, M. (2022). An Intelligent Cognitive-Inspired Computing with Big Data Analytics Framework for Sentiment Analysis and Classification. *Information Processing & Management*, *59*(1), 102758. https://doi.org/10.1016/j.ipm.2021.102758

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 02 (2024)

18. D. Tang, F. Wei, B. Qin, N. Yang, T. Liu and M. Zhou, "Sentiment Embeddings with Applications to Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 2, pp. 496-509, 1 Feb. 2016, doi: 10.1109/TKDE.2015.2489653.

19. Sosa, P. M. (2017). Neural Network Models for Sentiment Analysis on Twitter Data. *Proceedings of the IEEE International Conference on Big Data*.

20. Akana Chandra Mouli Venkata Srinivas et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1074 012007