International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 12th June 2022.
Link : https://ijiemr.org/downloads/Volume-11/Issue-06

Title: Twitter Sentimental Analysis Using Recurrent Neural Network and LSTM

volume 11, Issue 06, Pages: 1424-1437

Paper Authors: **Ch. B. V. Durga, V. Tanuja, J. V. A. Sowmya, M. SowmyaPoojitha, A.Siva Parvathi**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Twitter Sentimental Analysis Using Recurrent Neural Network and LSTM

Ch. B. V. Durga[1], V. Tanuja[2], J. V. A. Sowmya[3], M. SowmyaPoojitha[4], A. Siva Parvathi[5]

[1]Assisstant Professor CSE, PSCMR College of Engineering & Technology, Vijayawada, Andhra Pradesh

[2,3,4,5] Student, CSE, PSCMRCET, Vijayawada, Andhra Pradesh

ABSTRACT: -

As we already know, people use social media for a long time every day. So, we are taken Twitter data to predict whether the tweet was positive or negative. We are using an RNN& LSTM as ML algorithms for the prediction tweets. Twitter is a microblogging site for tweets to friends as a posting on their Twitter accounts. The tweets all are in text format so we are doing the text processing for those tweets. Text processing playsan important role in tweets as negative or non-negative. Here we are analyzing all sentiments.This project was based on airline tweets and used data from the Kaggle source. We are taken train data as 80% and test data as 20%. We got an accuracy of 83% for the proposed algorithm. To improve the performance, we are using a Bi-directional LSTM model. In this paper, we are analyzing the tweets on US Airlines data with two attributespositive and negative. We used both RNN & LSTM for getting accuracy. We are using programming for getting output is python.

Keywords: -

Sentimental analysis, tweets, Airline, Recurrent Neural Networks, LSTM (Long-Short-Term-Memory)

INTRODUCTION: -

We gathered datasets comprising customers' emotions and feelings about the services they utilized for sentiment analysis.Twitter is a social media platform where people may share their feelings and emotions through tweets.However, attempting to gain a basic understanding of this unstructured information (tweets) can be time-consuming. This material has no structure (assessments) on a specific site and is seen by the users, who then build an opinion about the voices or administrations, and finally form a specific judgment. These emotions are then summed up to amass criticism for various objectives in order to provide useful assessments where assumption investigation is used.

OnTwitter, we have a tweet of 140 characters. Tweets may be positive or negative. This is used for the user to choose better airline service for the journey to any place. This is nothing but a review of that airline's service.It presents numerous challenging opportunities for application development, owing to the massive growth of publicly available data on internet sources such as websites and interpersonalgroups.In sentiment analysis, we must make abinary conclusion based on the tweet or review, such as yes/no, like/dislike, or positive/negative. Opinion mining and emotion mining are two terms for the same procedure. Organizations in the advertising business utilize it to improve their strategies, to understand customers' sentiments toward products or brands, how people react to content deliveries.

It is used in the political sphere to track political ideas, discover consistency and inconsistency among explanations and activities at the administrative level, and also to predict the outcomes of political decisions. Sentiment analysis is useful in airline tweet data for recognizing user thoughts and making improvements to services based on customer reactions. The process of assigning at least one predetermined class to message records is known as text mining/classification. The terms "archives" and "tweets" both hint at a comparison idea in the current investigation. To illustrate an adaptable portrayal of the issue, we can consider each tweet as an archiveand apply text-classification principles like tokenization, stemming, term-recurrence, and record recurrence.

RELATED WORK: -

MeghaRathi [1] (et. al) got the highest accuracy for the hybrid decision tree. Three algorithms it is taking more time for calculating accuracy or sentimental analysis. They used a technique of NLP (Natural Language processing) for text mining. We can use an RNN (Recurrent Neural Network) algorithm for getting more accuracy.

ShihabElbagir (et.al) [2]Multinomial logistic regression, Support Vector Regression (SVR), and Random Forest (RF) are some of the known approaches. Finally, in this base paper, we are using the supervised algorithm of decision tree based on ordinal regression of Twitter sentimental analysis.

Shaunak Joshi (et.al) [3] In this paper, the existing algorithms are the Naïve Bayes algorithm, Entropy, and Baseline. NLP (Natural Language Processing) is the proposed algorithm. The data set contains both positive and negative tweets.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Meylan Wong Akar (et.al) [4], In this base paper, the existing algorithms are KNN (K-Nearest Neighbour) and SVM algorithms. The proposed algorithm is the Naive Bayes Algorithm. Here there are taken the tweet to predict whether the tweet they kept as positively or negatively. These are kept based on their emotion.

AdyanMarendaRamadhani (et.al) [5] In this base paper, we are using a deep learning method. They compared the tweets of both Korean and English tweets. The proposed algorithm is DNN (Deep Neural Network). The first hidden layer filters the words, the second filters based on the sentence, and the third are based on the popularity of the words based on the online dictionary. This paper is using an MLP algorithm. Better accuracy of using the DNN algorithm

3.PROPOSED SYSTEM: -

In this, we are using some python libraries to do coding for the twitter sentimental analysis. In this, we are doing four steps for getting output as a negative or non-negative. To analyse the data which are taken we are using some steps: -

1. Input Data
2. Pre-processing of data
3. Extraction of Characteristics
4. The Process of classification
5. Long Short-Term Memory
6. Activation Function

Input Data:

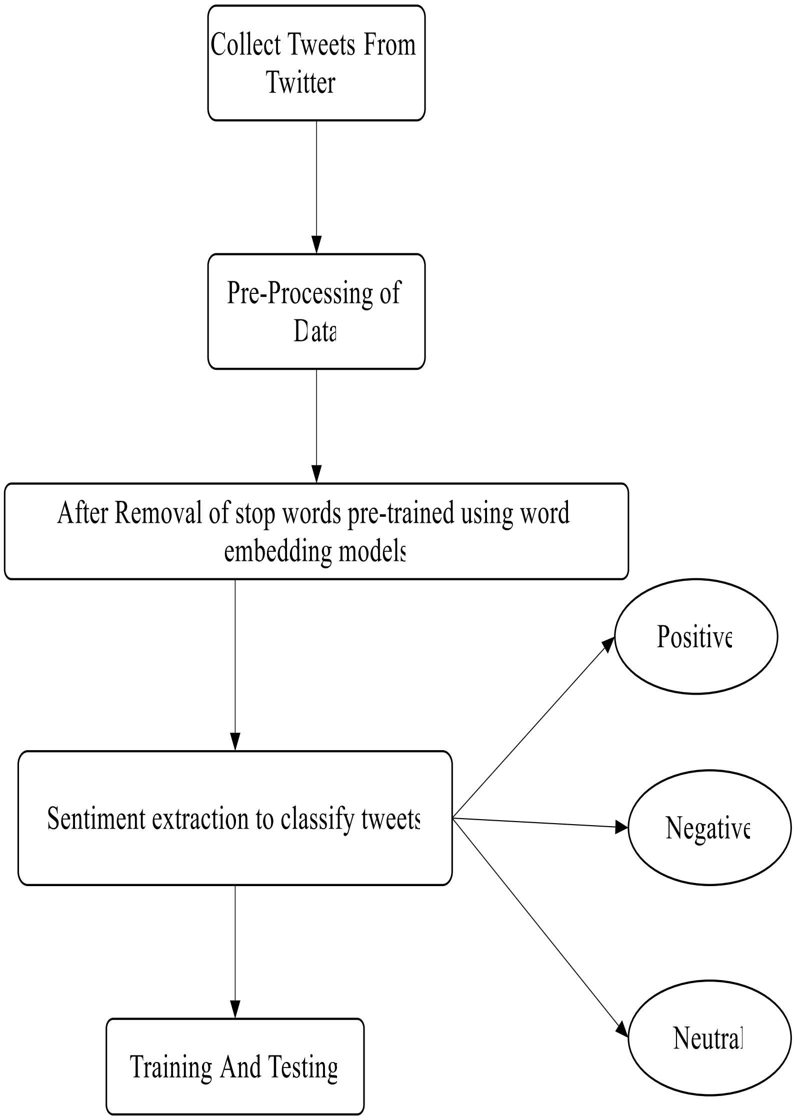
The input data is taken from the Kaggle repository. The Kaggle input data is sent to the input for data pre-processing.

DATA PRE-PROCESSING

The data is taken for input to do data pre-processing. In this data pre-processing it will be done for the data which are taken data a input. Here we are doing data pre-processing for input data. In this, we are doing

1. Remove Unicode strings from the tweets
2. Remove URL address
3. Remove the hashtag from the beginning of a term.
4. Get rid of the integers
5. Get rid of specific HTML entities (e.g., &)
6. Remove hyperlinks
7. Remove Punctuation and split (´s, ´t, ´ve) with space for filter
8. Remove whitespaces
9. Remove single space remaining at the front of the tweet
10. Remove usernames
11. Remove special characters
12. Removal of stop words

This all can be done in this data pre-processing. After completion of this data pre-processing the pre-processed data is taken as input for the feature extraction. We are using a coding language for data pre-processing in python.The schematic representation of internally doing work for the pre-processing is



1.1 Pre-processing

FEATURE EXTRACTION: -

This is the next immediate process after the data pre-processing. It is nothing but Data Cleaning. When the data is pre-processed, that data is input to the feature extraction. Then it will be getting a new feature that is suitable for that data. Then the feature is extracted for that data. Then it is moving to the further process. We achieved 83° of healing using the classic LSTM and RNN, but expect a unit specifically designed for our task to be able to extract feel better.We experimented with many architectural decisions but kept the recurring neural network and RNN algorithm works on algorithm. We applied RNN algorithm on dataset which is taken base cell design that already exists in the literature.

CLASSIFICATION PROCESS: -

Here we are a classification method to classify the tweets. To classify the tweets, we are using the Machine Learning algorithm RNN (Recurrent Neural Network).

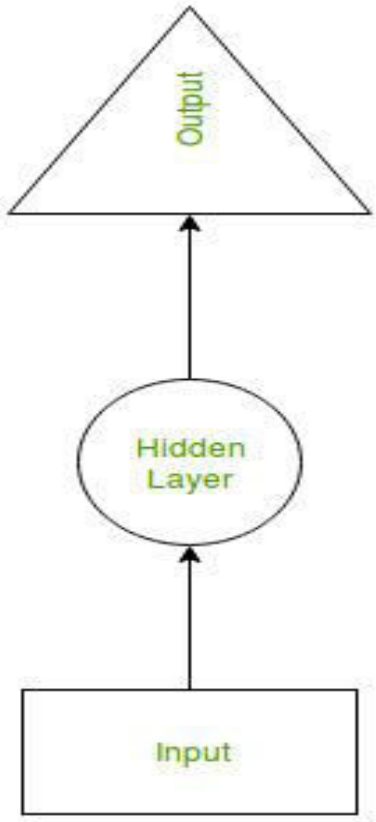
IMPLEMENTATION: -

Here we are using two types of methods to predict the tweets as negative or non-negative.

RecurrentNeural Network&Long Short-Term Memory:

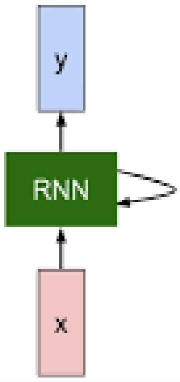
RNN is a deep learning model that may be used to analyze time series and text data. RNN is a form of deep learning approach that has the ability to remember data from the past.However, RNN has a difficulty known as the evanescent gradient problem.The weight update values are unaffected by the gradient value, indicating that this is a bug.

from the Kaggele. The RNN algorithm works on our data is splitting into certain layers and then the algorithm is working on the data. The working of RNN algorithm in iagramatic representation is



The current state of the RNN algorithm can be calculated using the following formula

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



$$h_t = f_W(h_{t-1}, x_t)$$

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$

where

h_t-> current state

h_{t-1} -> previous state

x_t-> input state

The formula for applying activation function is

Where

W_{hh} ->weight at the recurrent neuron

M_{xh} ->weight at input neuron

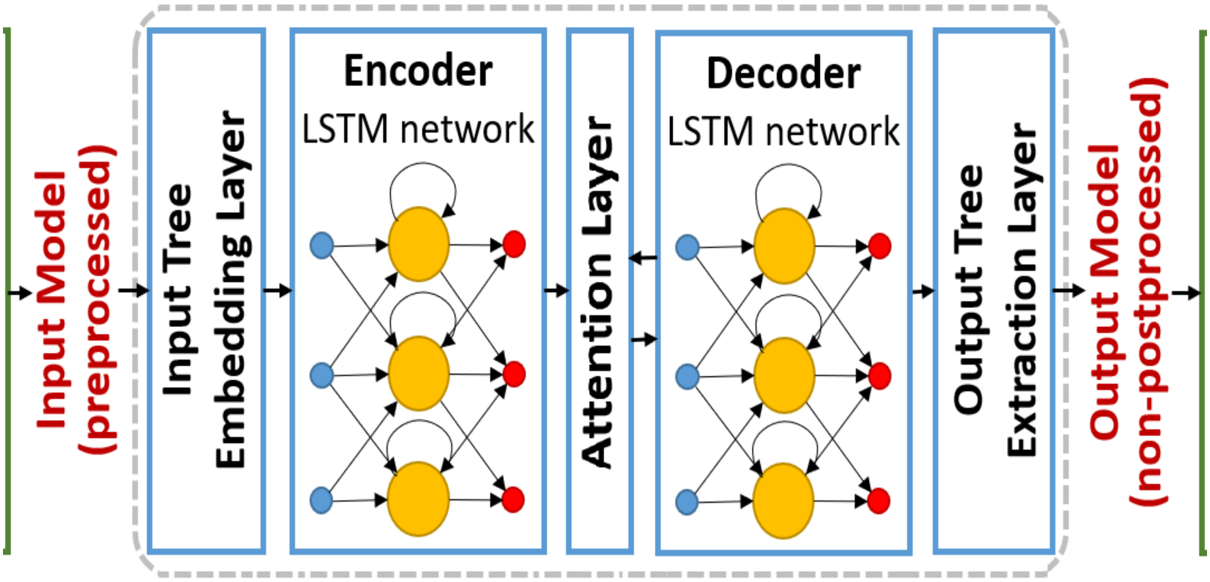
Formal calculation output is

Y_t -> output

W_{hy} -> weight of output layer

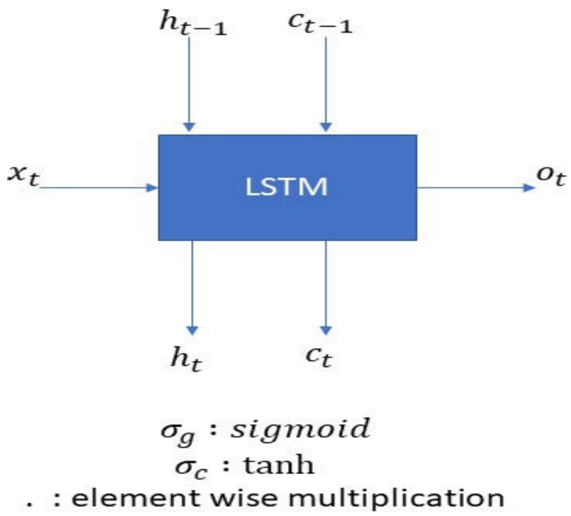
LSTM (Long Short-Term Memory): -

SeppHochreiter and JurgenSchmidhuber were the first to introduce LSTM. We're using the LSTM algorithm since our dataset has more data and time columns than the LSTM method can handle.The working of LSTM based on the Neural networks is



1.2 LSTM based on Neural Networks

Firstly, we are pre-processing the data then we are doing the feature extraction then we are applying the algorithm Of RNN and LSTM that is explained in the above figure.LSTMs were created to overcome this problem. The gate that can keep deleting data based on the situation has been forgotten by the LSTM. It features input and output gates as well. In general, the LSTM model outperforms the Simple RNN model. We used several text pre-processing algorithms on our dataset before using RNN or LSTM.

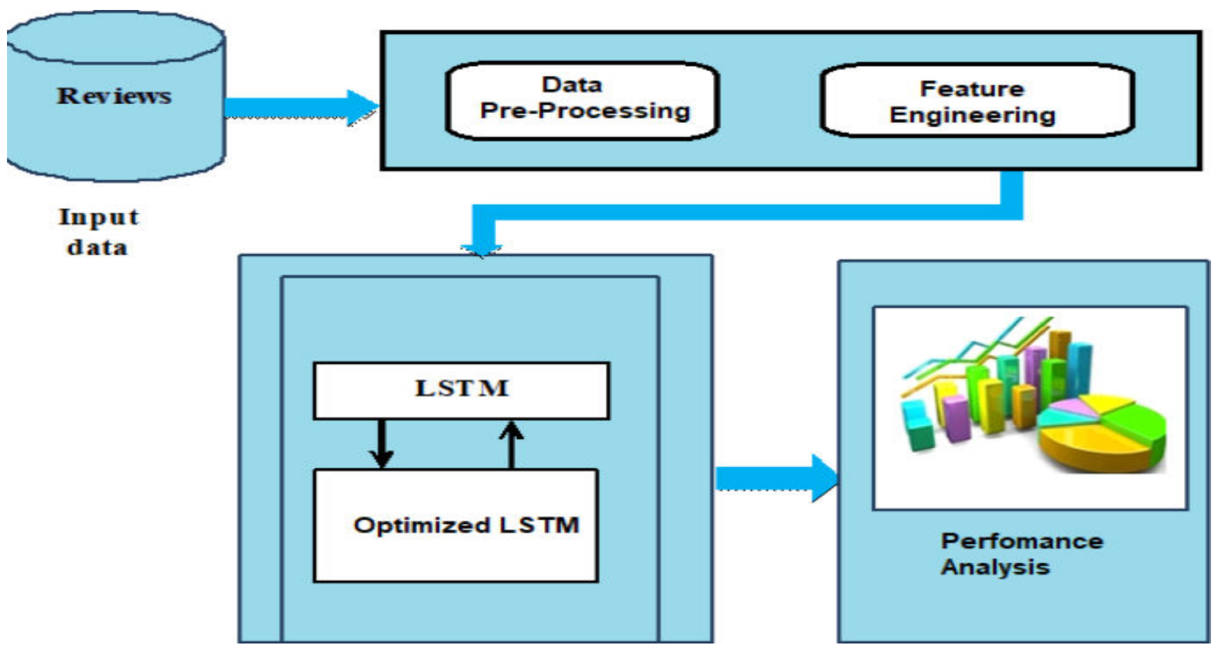


The formula for LSTM to pass the gate is

For the forward pass of an LSTM cell with a forget gate, the compact representations of the equations are:

$$f_t = \sigma_g\,(W_f \times x_t + U_f \times h_{t-1} + b_f)$$
$$i_t = \sigma_g\,(W_i \times x_t + U_i \times h_{t-1} + b_i)$$
$$o_t = \sigma_g\,(W_o \times x_t + U_o \times h_{t-1} + b_o)$$
$$c'_t = \sigma_c\,(W_c \times x_t + U_c \times h_{t-1} + b_c)$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t$$
$$h_t = o_t \cdot \sigma_c(c_t)$$

f_t *is the forget gate*
i_t *is the input gate*
o_t *is the output gate*
c_t *is the cell state*
h_t *is the hidden state*



1.4 Performance of LSTM

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

RESULT: -

We are using the precision, score, for getting a accuracy for all the these two types. We are set the epochs as 10. For.every epoch value is either be a increase or decrease the value of the epoch. So we got a accuracy of 83% for both the RNN and LSTM algorithm. We are used pandas library in spyder notebook. Our dataset contains approximately 1400 tweets. The columns are nearly 6. The prediction value and pre-

```
*Python 3.6.4 Shell*
File  Edit  Shell  Debug  Options  Window  Help
Python 3.6.4 (v3.6.4:d48eceb, Dec 19 2017, 06:54:40) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
=========== RESTART: C:\Users\TANUJA\OneDrive\Desktop\code\RNN.py ===========
Unique words:  16727

Tokenized review:
 [[57, 213]]
[[  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0 446]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0]]
                        Feature Shapes:
Train set:              (11200, 30)
Validation set:         (1400, 30)
Test set:               (1400, 30)
Sample input size:  torch.Size([50, 30])
Sample input:
 tensor([[   0,    0,    0,  ...,  541, 1048, 2576],
        [   0,    0,    0,  ...,  379,   59, 1331],
        [   0,    0,    0,  ..., 3467,    9,   63],
        ...,
        [   0,    0,    0,  ...,    7,    2,  458],
        [   0,    0,    0,  ...,   48,  176,  288],
        [   0,    0,    0,  ...,  216,  956, 8481]], dtype=torch.int32)

Sample label size:  torch.Size([50])
Sample label:
 tensor([1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0,
        1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1,
        1, 0], dtype=torch.int32)
No GPU available, training on CPU.
SentimentRNN(
  (embedding): Embedding(16728, 200)
  (lstm): LSTM(200, 128, num_layers=2, batch_first=True, dropout=0.5)
  (dropout): Dropout(p=0.3, inplace=False)
  (fc): Linear(in_features=128, out_features=1, bias=True)
  (sig): Sigmoid()
)
```

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

rounding value of the tweet is 0.003.

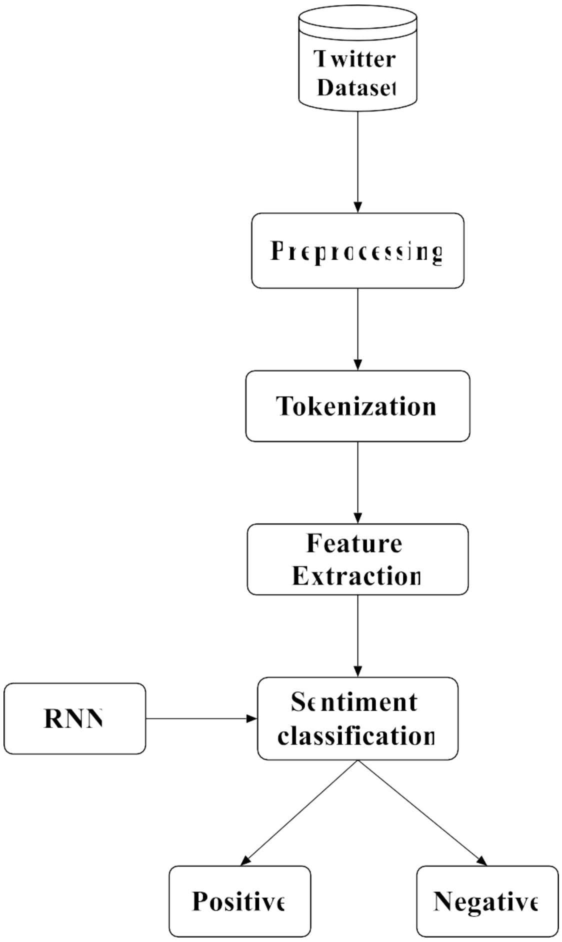
The epoch value of our tweets is: -

The screenshot of the output: -

```
Epoch: 1/10... Step: 100... Loss: 0.540416... Val Loss: 0.531690
Epoch: 1/10... Step: 200... Loss: 0.455831... Val Loss: 0.462834
Epoch: 2/10... Step: 300... Loss: 0.368951... Val Loss: 0.440977
Epoch: 2/10... Step: 400... Loss: 0.280798... Val Loss: 0.456464
Epoch: 3/10... Step: 500... Loss: 0.147571... Val Loss: 0.449907
Epoch: 3/10... Step: 600... Loss: 0.279145... Val Loss: 0.466536
Epoch: 4/10... Step: 700... Loss: 0.071539... Val Loss: 0.490733
Epoch: 4/10... Step: 800... Loss: 0.097295... Val Loss: 0.515668
Epoch: 5/10... Step: 900... Loss: 0.158322... Val Loss: 0.514359
Epoch: 5/10... Step: 1000... Loss: 0.157207... Val Loss: 0.642546
Epoch: 5/10... Step: 1100... Loss: 0.070588... Val Loss: 0.701057
Epoch: 6/10... Step: 1200... Loss: 0.058637... Val Loss: 0.834404
Epoch: 6/10... Step: 1300... Loss: 0.089804... Val Loss: 0.757070
Epoch: 7/10... Step: 1400... Loss: 0.039029... Val Loss: 0.714030
Epoch: 7/10... Step: 1500... Loss: 0.072273... Val Loss: 0.866350
Epoch: 8/10... Step: 1600... Loss: 0.009348... Val Loss: 0.957810
Epoch: 8/10... Step: 1700... Loss: 0.044470... Val Loss: 0.967256
Epoch: 9/10... Step: 1800... Loss: 0.012711... Val Loss: 0.762242
Epoch: 9/10... Step: 1900... Loss: 0.032697... Val Loss: 0.869017
Epoch: 9/10... Step: 2000... Loss: 0.010702... Val Loss: 0.884627
Epoch: 10/10... Step: 2100... Loss: 0.012138... Val Loss: 0.907627
Epoch: 10/10... Step: 2200... Loss: 0.015311... Val Loss: 0.958358
Test accuracy: 0.836
Test loss:
0.16428571428571426
[[5, 22, 11, 367, 5, 126, 11, 8, 10, 85, 335, 21, 922, 93, 194, 1550, 44, 3, 34, 125, 11, 2888]]
[[   0    0    0    0    0    0    0    0    5   22   11  367    5  126
    11    8   10   85  335   21  922   93  194 1550   44    3   34  125
    11 2888]]
1
Prediction value, pre-rounding: 0.003544
Negative review detected.
>>>
```

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

PROPOSED SYSTEM ARCHITECTURE: -



1.2 System Architecture

CONCLUSION: -

Here we are using deep learning techniques as an LSTM network model to analyze the tweets of the airline services. The tweets on social media are based on reviews of airline services.The tweets which are taken as a training dataset show better performance than the test data. We are taken train data and test data to predict the tweets as negative or non-negative.For the train data we are applying the pre-processing and then feature extraction, after that we are applying the classification process. We are applying an algorithm we are getting an accuracy of 83%.Furthermore, accuracy can be enhanced using the Bidirectional LSTM network (Bi-LSTM). wegot a better accuracy by applying both RNN

(Recurrent Neural Network) and LSTM (Long Short-Term Memory).

REFERENCES: -

1. Oduwa Edo-Osagie, Gillian Smith, Iain Lake, ObagheEdeghere, Beatriz De La Iglesia, "Twitter mining using semi-supervised classification for relevance filtering in syndromic surveillance", PLOS ONEhttps://doi.org/10.1371/journal.pone.0210689 July 18, 2019
2. Imane El Alaoui, Youssef Gahi, RochdiMessoussi, YounessChaabi, Alexis Todoskoff and AbdessamadKobi, "A novel adaptable approach for sentiment analysis on big social data", Journal of Bigdata, SpringerOpen,https://doi.org/10.1186/s40537-018-0120-0,2018.
3. MeghaRathi, Aditya Malik, DakshVarshney, Rachita Sharma, SarthakMendiratta "An overview: Sentiment Analysis of Tweets using Machine Learning Approach" August 2018| Jaypee Institute of Information Technology JIIT Sec-62.
4. V.Lakshmi, K.Harika, H.Bavishya, Ch. Sri Harsha, "SENTIMENT ANALYSIS OF TWITTER DATA," vol.04, February 2017.
5. ShihabElbagir 1,2 and Jing Yang1 "An overview: Twitter Sentiment Analysis based on ordinal Regression" The National Natural Science Foundation of China (Nos. 61672179, 61370083, 61402126).
6. P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," Proc. 10th Int. Work. Semant. Eval., pp. 1–18, 2016.
7. Shaunak Joshi, DeepaliDeshpande "An Overview: Twitter Sentiment Analysis System" International Journal of Computer Applications (0975 – 8887) Volume 180 – No.47, June 2018.
8. C Ramakrishna, G Kiran Kumar, A Mallikarjuna Reddy, and P. Ravi, "A Survey on various IoT Attacks and its Countermeasures," Int. J. Eng. Res. Comput. Sci. Eng., vol. 5, no. 4, pp. 2394–2320, 2018, [Online]. Available: http://ijercse.com/specissue/april-2018/27.pdf.
9. Swarajya Lakshmi V Papineni, SnigdhaYarlagadda, HaritaAkkineni, A. Mallikarjuna

Reddy. Big Data Analytics Applying the Fusion Approach of Multicriteria Decision Making with Deep Learning Algorithms International Journal of Engineering Trends and Technology, 69(1), 24-28, doi: 10.14445/22315381/IJETT-V69I1P204.

10. Meylan Wongkar1, Apriandy Angdresey2 "An Overview: Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter" Department of Informatics Engineering De La Salle Catholic University October 2019.8.
11. V.Lakshmi, K.Harika, H.Bavishya, Ch.SriHarsha, "SENTIMENT ANALYSIS OF TWITTER DATA," vol.04, February 2017. Link- https://www.irjet.net/archives/V4/i3/IRJET-V4I3581.pdf
12. Learning. 20 (3): 273–297. [5] Vladimir N. Vapnik. The Nature of Statistical Learning Theory. Springer, New York, 1995.
13. J.R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993 12. Predicting Student's Final GPA Using Decision Trees: A Case Study by Mashael A. Al-Barrak and Muna Al-Razgan
14. https://www.lexalytics.com/lexablog/2014/sentiment-analysis-added-tooxford-dictionaries
15. Mallikarjuna Reddy, A., Venkata Krishna, V. and Sumalatha, L. Face recognition approaches: A survey. International Journal of Engineering and Technology (UAE), 4.6,6(7)(2018) 117-121. doi: 10.14419/ijet.v7i4.6.20446.
16. M. R. Ayaluri, K. Sudheer Reddy, S. R. Konda, and S. R. Chidirala, "Efficient steganalysis using convolutional auto encoder network to ensure original image quality," PeerJ Computer Science, vol. 7, p. e356, 2021.
17. A Short Introduction to Boosting by Yoav Freund and Robert E. Schapire (AT&T Labs, Research, Shannon Laboratory)
18. R. Plutchick. "Emotions: A general psychoevolutionary theory." In K.R. Scherer & P. Ekman (Eds) Approaches to emotion. Hillsdale, NJ; Lawrence Erlbaum Associates, 1984.
19. P. Basile, V. Basile, M. Nissim, N. Novielli, V. Patti. "Sentiment Analysis of Microblogging Data". To appear in Encyclopedia of

Social Network Analysis and Mining, Springer. In press.

20. Johan Bollen, Huina Mao, and Alberto Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economicphenomena.," in International AAAI Conference on Weblogs and Social Media (ICWSM'11), 2011.
21. Cortes, Corinna; Vapnik, Vladimir N. (1995). "Supportvector networks". Machine Learning. 20 (3): 273–297.
22. Russell, Stuart; Norvig, Peter (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice-Hall. ISBN 978-0137903955.
23. Greene, William H. (2012). Econometric Analysis (Seventh ed.). Boston: Pearson Education. pp. 803–806. ISBN 978-0-273-75356-8.
24. W. Chu and S. S. Keerthi, "Support vector ordinal regression," Neural Comput., vol. 19, no. 3, pp. 792–815, 2007.
25. S. Liu, F. Li, F. Li, X. Cheng, and H. Shen, "Adaptive co-training SVM for sentiment classification on tweets," in Proceedings of the 22nd ACM international conference on Information & Knowledge Management, 2013, pp. 2079–2088.
26. J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986. [28] L. Breiman, "Random Forests, Vol. 45," Mach Learn, vol. 1, 2001.
27. Mallikarjuna Reddy A, RupaKinnera G, Chand rasekhara Reddy T, Vishnu Murthy G. 2019. Generating cancelable fingerprint template using triangular structures. Journal of Computational and Theoretical Nanoscience 16(5–6):1951-1955
28. A Mallikarjuna Reddy, VakulabharanamVenkata Krishna, LingamguntaSumalatha and AvukuObulesh, Age Classification Using Motif and Statistical Feature Derived On Gradient Facial Images, Recent Advances in Computer Science and Communications (2020) 13:965. https://doi.org/10.2174/2213275912666190417151247.
29. A. Celikyilmaz, D. Hakkani-Tür, and J. Feng, "Probabilistic model-based sentiment analysis of Twitter messages," in 2010 IEEE Spoken Language Technology Workshop, 2010, pp. 79–84.

30. A. Genkin, D. D. Lewis, and D. Madigan, "Large-scale Bayesian logistic regression for text categorization," Technometrics, vol. 49, no. 3, pp. 291–304, 2007.

31. A.Mallikarjuna, B. KarunaSree, " Security towards Flooding Attacks in Inter Domain Routing Object using Ad hoc Network" International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-3, February 2019.

32. B. Pruthvi Raj Goud, G. L. AnandBabu, G. Sekhar Reddy, A. Mallikarjuna Reddy "Multiple object detection interface using HSV, hough and haar-classifier" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019.