



SENTIMENT ANALYSIS USING DEEP LEARNING

B. Ramya¹, A. Jothi²

II M.SC CS¹, Assistant Professor², Department of Computer Science^{1,2},
K.R. College of Arts and Science^{1,2}.

ABSTRACT

Twitter is a well-known and widely used social networking site since it enables users to share their ideas and opinions on any topic as well as submit messages or comments from anywhere in the world. Techniques for Sentiment Analysis are employed to research and evaluate these reviews or opinions. Sentiment analysis is a NLP method that is used to specific opinions into exclusive sentiments like positive or negative. In this work, we take Airline Dataset from Twitter and did sentiment analysis on that dataset using LSTM. Binary text classifier to classify the sentiment behind the text and various NLPpreprocessing techniques to clean the data and the LSTM layers to build the text classifier. The dataset contains 14000 tweets data. This research work will show the sentiments with respect to the given sentences, given by the user.

Keywords- sentiment analysis, NLP, LSTM, twitter.

1. INTRODUCTION

Sentiment Analysis is procedure of amassing and inspecting information based totally upon the man or woman feelings, opinions and thoughts. Sentimental analysis often called as opinion mining as it mines the important feature from people opinions. Sentimental Analysis is done by using various deep learning techniques and Natural Language Processing (NLP) for the extraction of features from large data. Sentimental Analysis has various applications. It is used to generate opinions for people of social media by analyzing their feelings or thoughts which they provide in form of text. Sentimental Analysis is used in many real life scenarios, to get reviews about any product or movies, to get the financial report of any company, for predictions or marketing. Twitter is a micro blogging platform where anyone can read or write short form of message which is called tweets. This data is unstructured and written in natural language. Twitter Sentimental Analysis is the process of accessing tweets for a particular topic and predicts the sentiment of these tweets as positive or negative with the help of LSTM algorithm.

A part of machine learning is called deep learning. It contains millions of data points. It is a field that is focused on computer algorithms learning and developing on their own. Image categorization, language translation, and speech focus have all benefited from deep learning. It can tackle any pattern recognition problem without the need for human intervention. To provide correct results, deep learning algorithms require a significant amount of data, as just a result; information is fed as big data sets.

To evaluate whether data is favorable or negative, Natural Language Processing techniques like sentiment analysis is applied. NLP techniques are based on machine learning. NLP is specially used to understand the human language by machines, and translate human languages to other languages. In NLP contains some algorithms, has some databases of words that contains meanings of word, full forms of words and spelling correction. Because of NLP sentiment analysis that helps to analyze sentiment of tweets or any input given by human.

Binary Text Classifier to that classifies the sentiment behind the text. Classified unstructured text data into various categories. Binary classifiers are used when there are only two feasible solutions to a classification problem. We created a binary text classifier that divides tweet sentiment into positive

LSTM stands for Long Short Term Memory. We will build a neural network model to solve a basic sentiment analysis problem. The LSTM algorithm is used to build a model for classifying sentiment analysis. They are a particular RNN (Recurrent Neural Network) type that works well for problems involving sequence prediction. We can classify feedback based on emotion in a number of ways, but

we're using LSTM networks, which is the most recent technique. Using this method, the model can predict sentiment analysis on text, and predict the label.

2. RELATED WORK

Dr. G. S. N. Murthy, et al. [1] implemented LSTM networks are well-suited to classifying, processing, and making predictions supported statistic data since there may be lags of unknown duration between important events in a very statistic. The IMDB benchmark dataset is used for our experimental studies that contain movie reviews that are classified as being positive or negative.

Pedro M. Sosa [2] have propose 2 different Neural Network models that aim to combine the popular Long-Short Term Memory Neural Networks (LSTMs) with Convolutional Neural Networks (CNNs), and compare their accuracy against regular CNN and LSTM networks. Use Neural Networks to classify any given tweet into a “positive” or “negative” one.

Deb Dutta Das, et al. [3] by displaying the outcomes of a machine learning algorithm using R and Rapid Miner, we are classifying the sentiment of Twitter posts. The tweets are pre-processed, extracted, and then categorised into neutral, negative, and positive attitudes before the results are summarised overall. The Naive Bayes algorithm was used to categorise the attitudes expressed in recent tweets from the various airlines.

Maha heikal*, [4] suggested an ensemble model that combines Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models in order to predict the sentiment of Arabic tweets. On the Arabic Sentiment Tweets Dataset, our model exceeds the most recent deep learning model, which obtains an F1-score of 53.6% with a score of 64.46%. (ASTD).

Gurpreet Kaur, et al. [5] utilising SVM, Naive Bayes, and Random Forest as machine learning algorithms. Random forest is a supervised learning for Classification that works by building a huge number of Decision trees at training time and the output is dependent on the class mode of each tree. SVM algorithm used to classify the text into two classes as Positive and Negative. This is validated and verified by the Naïve Bayes algorithm.

M. Murali, et al. [6] the adaptive particle swarm optimization approach optimises weight parameters to enhance the performance of the LSTM. Extensive testing on four datasets shown that our suggested APSO-LSTM model outperformed more established techniques as standard LSTM, ANN, and SVM in terms of accuracy. The suggested model outperforms other current models in several measures, according to simulation findings.

Shilpa P C, et al. [7] This paper focus on classifying user emotion in Twitter messages using deep learning models such as LSTM (Long short Term Memory) and RNN (Recurrent Neural network). The feature extracted were combined and fed into an LSTM and RNN to train and predict the emotion classification labels of the tweet into negative or positive sentiments. The LSTM technique has a better capability of predicting sentiment classification accuracy.

Rida Khan, et al. [8] Sentiment Analysis is carried out using TextBlob analyzer. Consumer loyalty prediction is performed using Twitter data. Three classifiers are employed, namely, Random Forest, Decision Tree and Logistic Regression. A maximum classification accuracy of 99.05% is observed for Random Forest on 10-fold cross validation.

Peng Cen¹, et al. [9] used three deep learning networks (CNN, RNN, LSTM) to classify movie reviews from the IMDB dataset into positive and negative categories. Models that are often used for sentiment analysis are RNN and LSTM, CNN is often used in image recognition.

Dirash A R, et al. [10] the training dataset and test set are separated from the dataset. To resolve a fundamental sentiment analysis issue, we will construct a neural network model. The LSTM technique is employed to create a classification model for sentiment analysis. Long short term memory is referred to as LSTM. They are a particular RNN (Recurrent Neural Network) type that works well for problems involving sequence prediction.

Rakesh Suryawanshi, et al. [11] using Natural Language Processing Toolkit and techniques to determine sentiment of tweet where it is positive, neutral and negative. We use “Us airline dataset” which contains tweets about US airlines. . In lexicon based approach sentiment of word in database and predicted sentiment of word both are matched. We have used the VADER sentiment analysis. VADER is stand for Valence Aware Dictionary and Sentiment Reasoned it's a sentiment analysis tool that is rule based and lexicon based analysis tool. This tool helps to provide sentiment of tweets.

Nafees Akhter Farooqui, et al. [12] TextBlob has basic natural-language processing features. That is used to calculate the partiality and measure the polarity of tweets. In this study, we develop a model to categorise opinions on Twitter and other popular blogging platforms into three categories: favorable, negative, and neutral.

Anurag Lahon, et al. [13] classified twitter data into feedbacks using machine learning and natural language processing (NLP) with the help of programming language Python. The tweets are categorized into neutral, positive, negative sentiments. We will use different machine learning algorithms.

Ch. Satyanarayana, et al. [14] applying a deep learning technique to perform Twitter sentiment analysis. Simple Neural Network, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) methods are applied for the sentiment analysis and their performances are evaluated. The LSTM is the best among all proposed techniques with the highest accuracy.

3. DATASET

Twitter US Airline dataset consist of 14,000 tweets from the kaggle. This dataset is, contains tweets on US Airline of February 2015 classified in positive, negative and neutral tweets. It contains tweets on six major United States (US) airlines. The negative tweets are also classified in base of the negative reason.

	text
0	@VirginAmerica What @dhepburn said.
1	@VirginAmerica plus you've added commercials t...
2	@VirginAmerica I didn't today... Must mean I n...
3	@VirginAmerica it's really aggressive to blast...
4	@VirginAmerica and it's a really big bad thing...
5	@VirginAmerica seriously would pay \$30 a fligh...
6	@VirginAmerica yes, nearly every time I fly VX...
7	@VirginAmerica Really missed a prime opportuni...
8	@virginamerica Well, I didn't...but NOW I DO! :-D
9	@VirginAmerica it was amazing, and arrived an ...
10	@VirginAmerica did you know that suicide is th...
11	@VirginAmerica I pretty graphics. so muc...
12	@VirginAmerica This is such a great deal! Alre...

Fig.1: Dataset Text

text	airline_sentiment
@VirginAmerica What @dhepburn said.	neutral
@VirginAmerica plus you've added commercials t...	positive
@VirginAmerica I didn't today... Must mean I n...	neutral
@VirginAmerica it's really aggressive to blast...	negative
@VirginAmerica and it's a really big bad thing...	negative
@VirginAmerica seriously would pay \$30 a fligh...	negative
@VirginAmerica yes, nearly every time I fly VX...	positive
@VirginAmerica Really missed a prime opportuni...	neutral
@virginamerica Well, I didn't...but NOW I DO! :-D	positive
@VirginAmerica it was amazing, and arrived an ...	positive
@VirginAmerica did you know that suicide is th...	neutral
@VirginAmerica I pretty graphics. so muc...	positive
@VirginAmerica This is such a great deal! Alre...	positive
@VirginAmerica @virginmedia I'm flying your #f...	positive
@VirginAmerica Thanks!	positive
@VirginAmerica SFO-PDX schedule is still MIA.	negative
@VirginAmerica So excited for my first cross c...	positive
@VirginAmerica I flew from NYC to SFO last we...	negative
I ❤️ flying @VirginAmerica. ☺️👍	positive

Fig.2: Tweets in Dataset

4. METHODOLOGY

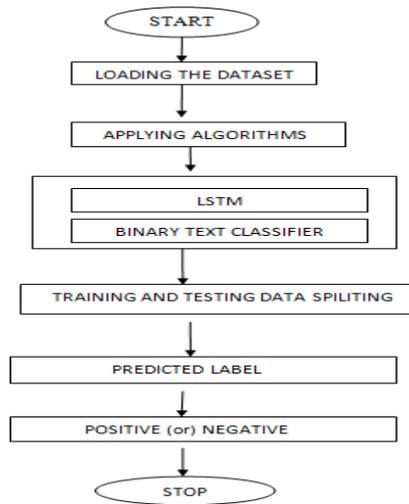


Fig. 3: Methodology of work

LSTM

LSTM stands for Long Short Term Memory. It is a particular RNN subtype that works well for challenges involving sequence prediction. It may anticipate the following word based on the text that came before. LSTM can carry out relevant information throughout the processing of inputs and it discards non-relevant information. LSTMs are designed to overcome the vanishing gradient problem and allow them to retain information for longer periods compared to traditional RNNs.

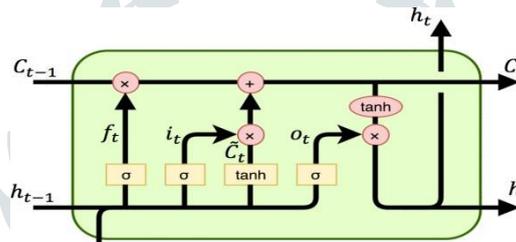


Fig.4: Basic Structure of LSTM

LSTM GATES

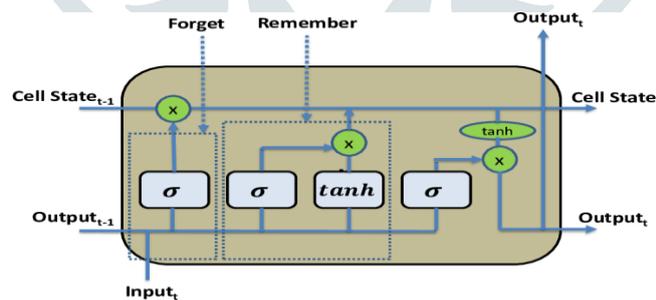


Fig.5: Forget gate, Input gate, Output gate

It has three important gates

- i. Forget gate
- ii. Input gate
- iii. Output gate

1. Forget gate

Forget gate is used to keep the important information and it discards the non relevant information. Previous hidden gate h_{t-1} and Current Event vector x_t are joined together $[h_{t-1}, x_t]$ and multiplied with the weight matrix W_f and passed through the sigmoid activation function with some bias to form Forget Factor f_t .

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

2. Input gate

Input gate is used to keep the current state of the information. Input gate is used to quantify the importance of the new information carried by the input.

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

$$c_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c)$$

The previous hidden gate (h_{t-1}) and Current Event (x_t) are joined together then multiplied with the weight matrix passed through the sigmoid activation function with some bias to produce it. Previous hidden gate (h_{t-1}) and Current Event (x_t) are joined together then multiplied with weight matrix passed through tanh activation function with some bias to produce c_t .

3. Output gate

It is used to decide what part of the current cell makes to output.

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

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

BINARY TEXT CLASSIFIER

In the challenge of binary classification, we must categorize our observations into one of the two labels based on the features. Binary Text Classifier to that classifies the sentiment behind the text. Classified unstructured text data into various categories. We created a binary text classifier that divides tweet sentiment into positive and negative categories. Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1.

5. WORKING OF THE MODEL

These reviews and opinions are examined using Sentiment Analysis tools. A NLP approach called sentiment analysis is employed to categorize opinions into different feelings, such as positive or negative. Whether their clients have favorable or negative opinions of them, every industry that interacts with customers is interested in learning what those opinions are. We create a binary text classifier to categorize the text's emotional content. The data is cleaned using a variety of NLP preprocessing methods, and the text classifier is constructed using LSTM layers. Long short term memory (LSTM) was designed over simple RNNs for modeling temporal sequences and their long-range dependencies more accurately. We created a binary text classifier that divides tweet sentiment into positive and negative categories.

SENTIMENT ANALYSIS MODEL

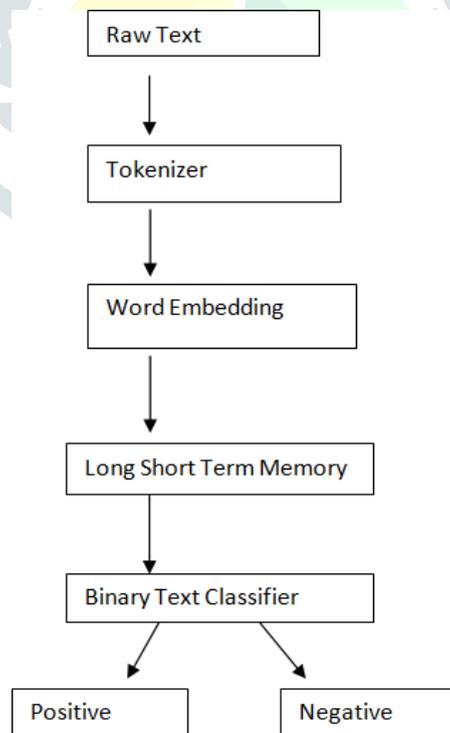


Fig.6: Sentiment Analysis Model

TOKENIZER

Tokenization is a straightforward procedure that turns raw data into a meaningful data string. Tokenization is a technique used in natural language processing to break down phrases and paragraphs into simpler language-assignable elements. The collecting of data (a sentence) and its breakdown into comprehensible components are the first steps of the NLP process (words).

WORD EMBEDDING

Word embedding is essential to natural language processing. Each is a real number vector representation of a specific word. Contextual information about that word is encoded within the vector numbers. To accomplish this, we first assign an individual number to each of the distinctive words before replacing the word in question with the number.

LSTM

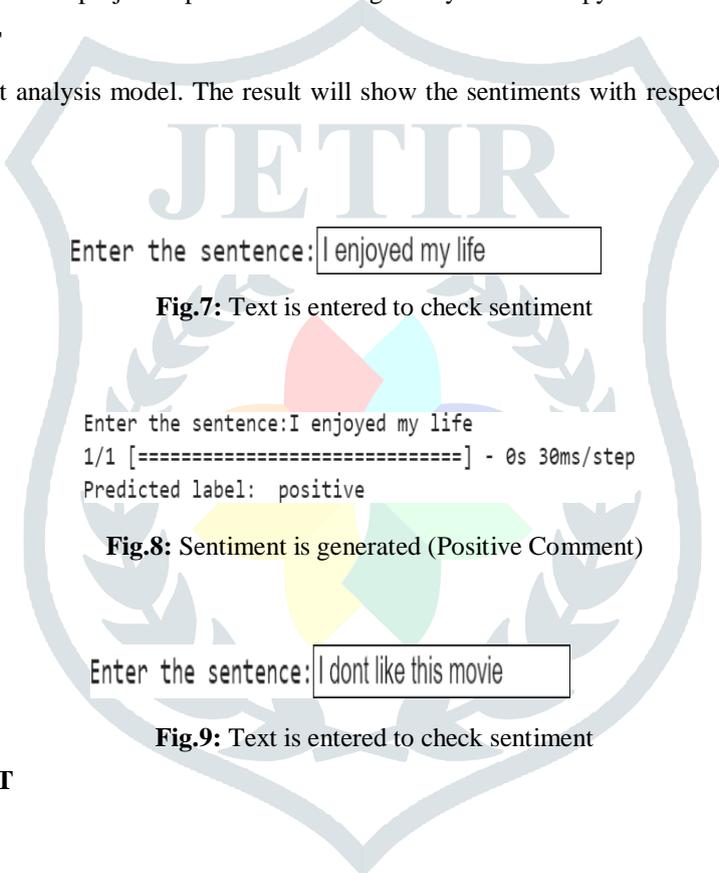
Our model's architecture is made up of three layers: an embedding layer, an LSTM layer, and a dense layer. We added the Dropout mechanism in-between the LSTM layers to prevent over fitting. On the other hand, LSTMs perform well with this type of content. Time-series data are highly suited for LSTM networks. One method of regularization is dropout. We built a binary text classifier that classifies the sentiment of the tweets into positive and negative. Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1.

BINARY TEXT CLASSIFIER

Binary Text Classifier to that classifies the sentiment behind the text. Classified unstructured text data into various categories. We created a binary text classifier that divides tweet sentiment into positive and negative categories. The actual class output, which can only be 0 or 1, is compared to each of the projected probabilities using binary cross entropy.

6. EXPERIMENTAL RESULT

Following are results of sentiment analysis model. The result will show the sentiments with respect to the given sentences, given by the user.

INPUT


Enter the sentence: I enjoyed my life

Fig.7: Text is entered to check sentiment**OUTPUT**

Enter the sentence:I enjoyed my life
1/1 [=====] - 0s 30ms/step
Predicted label: positive

Fig.8: Sentiment is generated (Positive Comment)**INPUT**

Enter the sentence: I dont like this movie

Fig.9: Text is entered to check sentiment**OUTPUT**

Enter the sentence:I dont like this movie
1/1 [=====] - 0s 26ms/step
Predicted label: negative

Fig.10: Sentiment is generated (Negative Comment)

7. CONCLUSION AND FUTURE SCOPE

Sentiment analysis is a NLP method that is used to convey opinions into different sentiments like positive or negative. Twitter Social Network site is used to read people's feelings, opinions, and views about any product or service. To read the opinions of the customer, we used sentiment analysis or opinion mining Natural Language Processing techniques. In this paper, we used deep learning algorithms to do sentiment analysis of Twitter dataset of the Airline system. Twitter Sentimental Analysis is the process of accessing tweets for a particular topic and predicts the sentiment of these tweets as positive or negative with the help of LSTM algorithm. In this work binary text classifier to classify the sentiment behind the text and various NLP preprocessing techniques to clean the data and the LSTM layers to build the text classifier.

In future we are planning to extend this study to a larger extent where different embedding models can be considered on large variety of the datasets.

REFERENCES

- [1] Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, Mounika Bagadi, Mounika Belusonti "Text based Sentiment Analysis using LSTM", International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 05, May-2020.
- [2] Pedro M. Sosa, "Twitter Sentiment Analysis using combined LSTM-CNN Models", Accelerating the world's research, June 7, 2017.
- [3] Deb Dutta Das, Sharan Sharma, Shubham Natani, Neelu Khare and Brijendra Singh, "Sentimental Analysis for Airline Twitter data", IOP Conf. Series: Materials Science and Engineering 263 (2017) 042067.
- [4] Qurat Tul Ain*, Mubashir Ali*, Amna Riaz†, Amna Noureen‡, Muhammad Kamran‡, Babar Hayat* and A. Rehman*, "Sentiment Analysis Using Deep Learning Techniques: A Review", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017.
- [5] Anurag Lahon, "Twitter Airline sentiment Analysis", IJEDR 2019 | Volume 7, Issue 3.
- [6] Gurpreet Kaur, Dr. Kamal Malik, "A Sentiment Analysis of Airline System Using Machine Learning Algorithms", International Journal of Advanced Research in Engineering and Technology (IJARET) Volume 12, Issue 1, January 2021.
- [7] J. Shobana1 · M. Murali1, "An efficient sentiment analysis methodology based on long short-term memory networks", Complex & Intelligent Systems (2021) 7:2485–2501.
- [8] Shilpa P C, Rissa Shereen, Susmi Jacob, Vinod P, "Sentiment Analysis Using Deep Learning", Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).
- [9] Rawaa Alatrash*, Hadi Ezaldeen, Rachita Misra, and Rojalina Priyadarshini, "Sentiment Analysis Using Deep Learning for Recommendation in E-Learning Domain".
- [10] Rida Khan, Siddhaling Urolagin, "Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction using Twitter Data", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018.
- [11] Peng Cen1, Kexin Zhang1 and Desheng Zheng1, "Sentiment Analysis Using Deep Learning", Approach Journal on Artificial Intelligence, vol.2, no.1, pp.17-27, 2020.
- [12] Dirash A R1, Dr. S K Manju Bargavi2, "LSTM Based Sentiment Analysis", International Journal of Trend in Scientific Research and Development (IJTSRD) Volume 5 Issue 4.
- [13] Ditiman Hazarika 1, Gopal Konwar 1, Shuvam Deb 1, Dr. Dibya Jyoti Bora 2, "Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing", Proceedings of the International Conference on Research, Vol. 24.
- [14] Rakesh Suryawanshi*1, Akshay Rajput*2, Parikshit Kokale *3, Prof. Subodh S. Karve*4, "Sentiment Analyzer Using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Volume:02/Issue:06/June -2020.
- [15] Nafees Akhter Farooqui, Ritika, Aayush Saini, "Sentiment Analysis of Twitter Accounts using Natural Language Processing", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-3, February 2019.