Sarcasm Detection Using Stacked Bi-Directional LSTM Model

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF DEGREE

OF

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted By

ASHISH PATIDAR (2K19/CSE/03)

Under the supervision of

DR. MANOJ KUMAR

(Associate Professor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

JUNE, 2021

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

DECLARATION

I, Ashish Patidar, Roll No. 2K19/CSE/03 student of M.Tech (Computer Science & Engineering), hereby declare that the Project Dissertation titled "**Sarcasm Detection Using Stacked Bi-Directional LSTM Model**" which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi. Report of the Major II which is being submitted to Delhi Technological University, Delhi, in partial fulfillment for the requirement of the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associate ship, Fellowship or other similar title or recognition.

Place: DTU, Delhi | Ashish Patidar

Date: 14-06-2021 | (2K19/CSE/03)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College Of engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I, hereby certify that the Project Dissertation titled *"**Sarcasm Detection Using Stacked Bi-Directional LSTM Model**"* which is submitted by Ashish Patidar, Roll No. 2K19/CSE/03, Department of computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment for the requirement of the award of degree of Master of Technology (Computer Science and Engineering) is a record of a project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date: 14/06/2021

Dr. Manoj Kumar
SUPERVISOR
Associate Professor
Dept. of Computer Science and Engineering
Delhi Technological University, Delhi

ACKNOWLEDGEMENT

I am most thankful to my family for constantly encouraging me and giving me unconditional support while pursuing this research.

I am extremely grateful to **Dr. Manoj Kumar** Associate Professor, Department of Computer Science Engineering, Delhi Technological University, Delhi for providing invaluable guidance and being a constant source of inspiration throughout my research. I will always be indebted to his for the extensive support and encouragement he provided.

I also convey my heartfelt gratitude to all the research scholars of the web Research Group at Delhi Technological University, for their valuable suggestions and helpful discussions throughout the course of this research work.

Ashish Patidar

Roll.No. 2K19/CSE/03

ABSTRACT

Sarcasm is one of the sentiments which is being used to communicate a negative opinion utilizing positive words. The world is full of social media and many kinds of the web-based portal and this media stores a huge amount of textual data which contains sentiments, and sarcasm is one of the sentiments which is being used nowadays in many of this platform, using sarcasm someone can communicate their negative words in a positive way which is we can call a sarcastic way of communication. In opinion mining, the field of natural language processing detection of sarcasm from a given data is an important task. It is a binary classification task for which model proposed a system which classifies whether a given set of word is sarcastic or not-sarcastic. In this research work, we proposed the work based on the Stacked Bi-Directional Long Short-Term Memory (Stk-BLSTM) network which enhances the overall result in terms of performance matrix.

Keywords—Sarcasm detection, Natural Language Processing, Stacked LSTM, Machine learning, Social data.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Sarcasm detection is a field of NLP (Natural Language Processing), a specific case of opinion mining where instead of detecting sentiment in the whole context, the focus is on sarcasm, therefore the task of this field in NLP is to detect whether a given sentence is sarcastic or not. The first problem that we have come across is that unlike in semantics and sentiment analysis tasks the sentiment categories are very well defined but the sharp corners of sarcasm are not that well defined. And it is an important task that before starting we have to detect it. "Sarcasm is a form of figurative language where the literal meaning of words cannot hold, and instead, the opposite interpretation is intended in a text" [1]. It is found out that 3960 Million people use social media which is 50.60% of the total world population and the average number of social media accounts is 8.1 per person so basically a huge amount of data is generating by this amount of social media users [2] now detecting the sarcastic comment might help to explore the new way of learning from social website content better and contains many applications such as sentiment analysis [3]. These social media now realized that there should be mitigation or restriction on the type of information to be posted by applying labeling on the information based on the level of sentiments and in this type of labeling there is a huge importance of sarcasm detection.

1.2 General

AI is slowly changing the conventional ways of computing. In the conventional programming paradigm, the problem is divided into sub-problems and then these sub-problems are assigned computing resources (programmers finding pattern and coding) for generating solutions. But, in the case of AI, an enormous volume of data of a problem is feed to a neural network that deduces patterns based on inputs to predict output with high accuracy and efficiency. In Fig 1.1 the AI field, deep learning is the most influential sub-field of machine learning which itself is a sub-domain of AI.

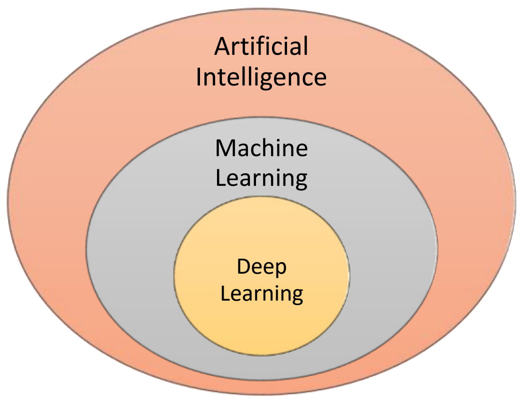


Fig 1.1 AI, machine learning and deep learning

Deep learning is being used to solve a variety of problems in various domains such as image pattern analysis. In AI, machine learning such as SVM, decision trees, etc. involves input, feature extraction, and classification for outputs. But in deep learning, feature extraction and classification are combined for predicting output (as shown in Fig. 2).

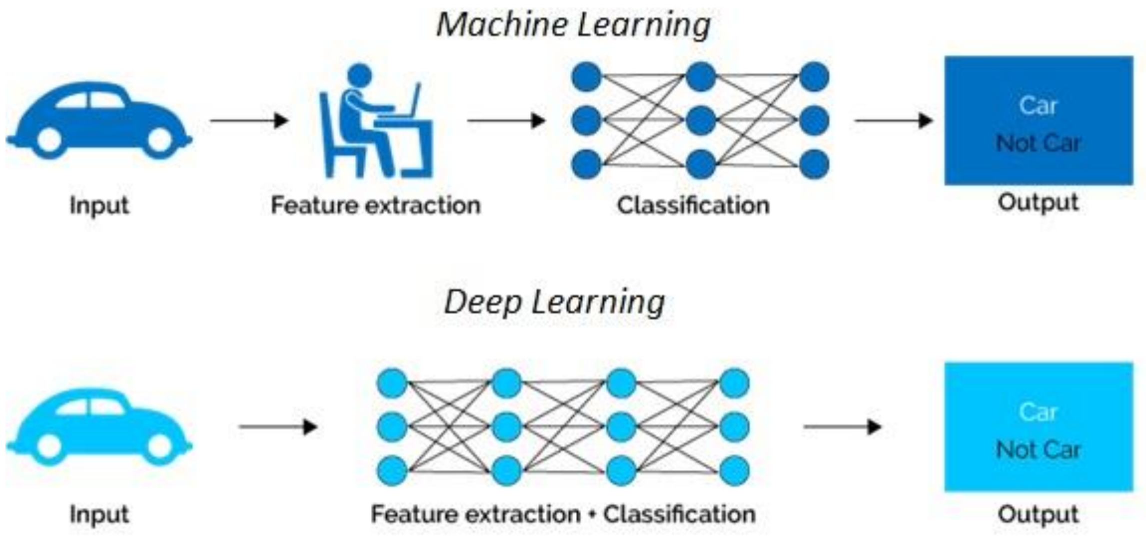


Fig 1.2 Machine learning vs Deep Learning

1.3 Introduction to Proposed Work

In recent years, [4] designed rules to which the system can identify the difference between a positive sentiment, negative sentiment, and also neutral sentiments for identifying sarcasm. In [5] the model detects sarcasm and builds semantic similarity and the role of effective featured content with the help of semantic resources. It is a challenging task owing to the lack of intonation and facial expressions in text. Nonetheless humans can still spot a feelings of person by their facial expressions.

The main focus is on social media as many people especially social activists make use of sarcasm in their tweets or post to express their point of view, and detecting the presence of sarcasm in a particular set of words is not an easy task for humans as well as machine. For example "It is a great feeling to wear face mask and specs together" it is containing sarcasm.

expressing negative sentiment about the situation of wearing the face mask and specs together using positive opinion words like ‘‘great feeling’’ this makes sarcasm detection an important task in the field of NLP to improve the overall performance to analyze sentiments effectively.

As part of this study, we proposed a deep learning model for sarcasm detection. The objective and working principle of this work can be elaborated as.

• In this proposed work first of all we preprocess the data using some machine learning algorithms like regression and clustering and also by filling the missing value manually by mean or most probable values or by deleting the whole tuple from the dataset.

• After preprocessing of data word embedding is used for the representation of words for analysis, In the form of a real-valued vector that decodes the meaning of the word such that the words that have the same or similar meaning in the vector space so that it is easy for our model to learn and predict the output with higher accuracy.

• In the proposed work we are presenting a Stacked-based Bidirectional Long Short-Term Memory (Stk-BiLSTM) which is being implemented using some machine learning library namely TensorFlow and Keras. The model concentrates on fetching the important words in terms of remarkable presence in a particular sentence or set of words, which is responsible to improve overall performance in the detection of sarcasm in a given data.

• We find out the result of the proposed model and then perform a comparative study of the model with two related and prior research work on the same dataset.

• Traning and Testing play an important role in any machine learning or deep learning model to improve the overall performance suitable amount of data is being used for training and training a deep learning model, Stk-BLSTM network.

• The proposed model is for classification problem so at the end you able to classify whether a given input comment is sarcastic or not-sarcastic.

Chapter 2

Related Work

2.1 Overview

In this section, we will synthesize our literature review. In the era of the internet many users use many kinds of social media and creating thousands GBs of data every day the internet is full of content and information, Twitter is one of the most famous and has a huge amount of active user with a good level of participation it has also emerged as a huge source of rich sentiment data [6]. Tseng et al. analyzed the text-based data and then able to find out the opinion on the basis of data of teaching staff by evaluating the result of the best outstanding teacher from the whole teaching staff with the help of attention-based LSTM [7]. In [8] author proposed a model by combining two methods and that is CNN and bidirectional gated recurrent unit for detecting a different kind of semantic in a particular sentence. In [9] author presented the lexicon-based enhanced LSTM model by using lexicon as pre-trained information which is further used as a classification problem. In [10] proposed a pattern-based method that used pattern recognition algorithms for detecting sarcasm on a Twitter dataset and discover four different types of sarcasm based on their set of features. In the next three proposed models, they have used bi-directional LSTM with some different deep learning techniques. [11] In this research work, the author proposed a deep learning model which makes use of soft attention in addition to the bi-directional LSTM model that is based on CNN which uses word embedding function with the help of global vectors word representation which encode the same meaning word for suitable sentiment analysis. In [12] author proposed the model with the help of bi-directional LSTM using multi-head attention for the detection of sarcastic comments. In this research work, the author proposed that this multi-head attention mechanism in addition to bidirectional LSTM enhances the overall performance, and it also performs better than Support Vector Machine models for detecting sarcasm in given context data. . In [13] author proposed a question answering (QA) network which is based on multiple dimension for sarcasm detection it also introduces the sufficient semantic information to understand the uncertainty in presence of sarcasm by using multiple dimension representation, with the help of deep QA answering network relies on the bi-LSTM model along with the feature of attention-based mechanism to find out the sarcasm in the conversation context-based information and also uses rule-based models which contain set of rules to detect the sarcasm in a given corpus. In [14] the author making use of google search to detect the sarcasm of smiles. [15] presented a model which mainly focuses on hashtags as a remarkable indicator but sometimes hashtags are not being used correctly especially in the case of sarcasm author has used the Twitter dataset for the training and testing purpose. [16] presented a model which is based on classification and make use of parse tree using phrase level model for detecting the sarcasm [4] uses a rule-based classifier to classify the positive phrase as well as the negative phrase, to detect the sarcasm.

2.2 Approaches Used for Sarcasm Detection

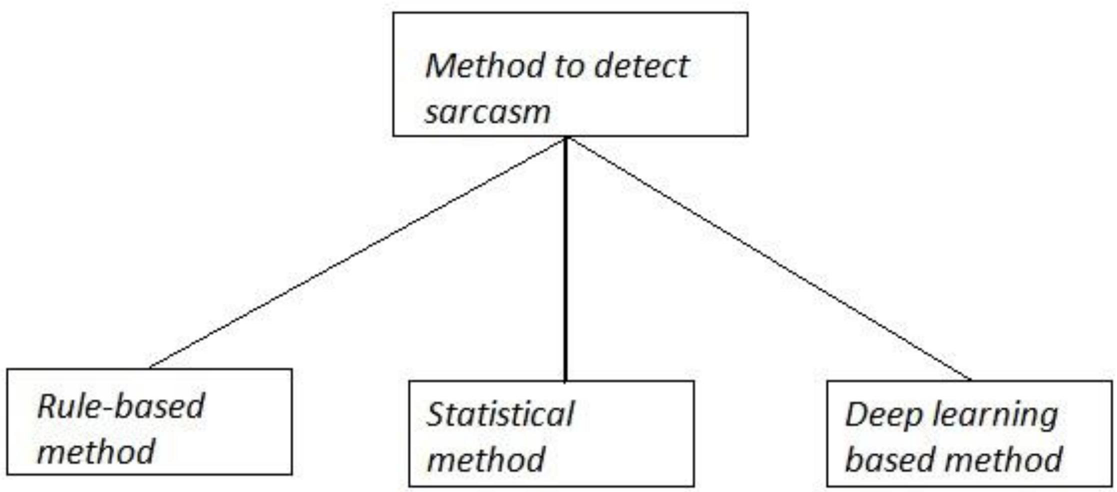


Fig 2.1 Approaches Used for Sarcasm Detection

2.1.1 Rule-based method

Rule-based techniques [17] detect sarcasm using particular evidence. These are some rules that aid in the detection of sarcasm. Various researchers have conducted various studies in order to obtain the rules. Some authors have spent time debating whether a specific smile is meant to be sarcastic or otherwise. The author conducted a Web search to determine the likelihood of a simile. They utilised nine different processes to verify the smile they were provided. A smile is recognized at each step using a set of rules. The advantage of utilising this method is that it provides an error analysis for numerous rules. Hashtags have been noted by many authors as a sign of sarcasm. Authors employ hash-tags like this to draw attention to sarcasm.

2.1.2 Statistical method

In terms of characteristics and learning algorithms, statistical techniques to sarcasm detection differ [18]. There are two varieties of these, which are explained below.

Features that have been used for sarcasm identification, various features are grouped. One of the elements of this method is a bag of words. In addition to these, it has a number of other traits that can be used to determine if a remark is sarcastic or not.

Various classifiers have been identified for detecting sarcasm. SVM is used by the majority of researchers. The Chi-squared test is also used to determine sarcastic characteristics. Sarcasm can also be detected using Naive Bayes. A comparison of rule-based versus SVM-based classifiers has been done by some authors.

2.1.2 Deep learning-based method

Deep learning techniques are becoming increasingly popular in today's society. Deep learning is one of the few strategies for detecting automatic sarcasm that is listed. One of the criteria used to detect sarcasm is the usage of similarity among word embedding. The use of congruent and incongruent word pairings has been shown to improve performance. This is done mostly on the basis of augmentation. Several researchers have combined convolutional neural networks and DNN and evaluated their findings to those of SVM. Deep learning has outperformed all other techniques in terms of results [19].

Chapter 3

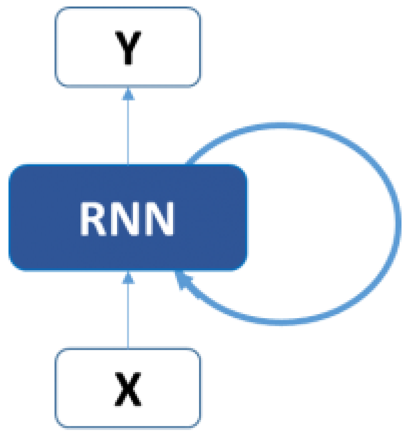
Theoretical Concepts

3.1 RNN (Recurrent Natural Network)

Artificial Neural Network is a advance computing system which is designed in such a way that it tries to work or analyze things as like human brain, which is the basic foundation of AI (Artificial Intelligence). If we take few example of sequential data, like housing price of few past years, for a particular location. A basic machine learning model or an AI based model may learn and to predict the housing price for the future ahead with the help of the pattern and features of past years data on the basis of size of house, construction quality and physical locality of the house.

In the traditional feed-forward neural network, all the result set are considered as independent. Which means if we are fitting model for a particular day, it will not consider the housing price of past days. So this dependency over time is achieved by RNN (Recurrent Neural Networks)

Input



Output

Fig 3.1 RNN (Recurrent Neural Networks)

RNNs can able to solve our problem of sequential data handling to a great extent but not up to the mark. Now a day we expect that our machines should be well enough to handle short contexts. But before going ahead we have to understand that our model should aware of context behind the sequential dataset, exactly like human brain. Which is not possible using this basic RNN model.

There are certain limitation of RNN. In order to add more information to the model, it converts the current information completely by using some function. And due to this, the overall information get modified, so that there is no differentiate between for important information and less important information.

On the other hand LSTM is modification over RNN, that collectively modify the information by addition and multiplication. In LSTM the flow of information is in such a way and that mechanism called as cell states. Using this the model selectively forget or remember things. The information at a specific cell state has three different sub-dependencies. These dependencies can be generalized to any problem as:

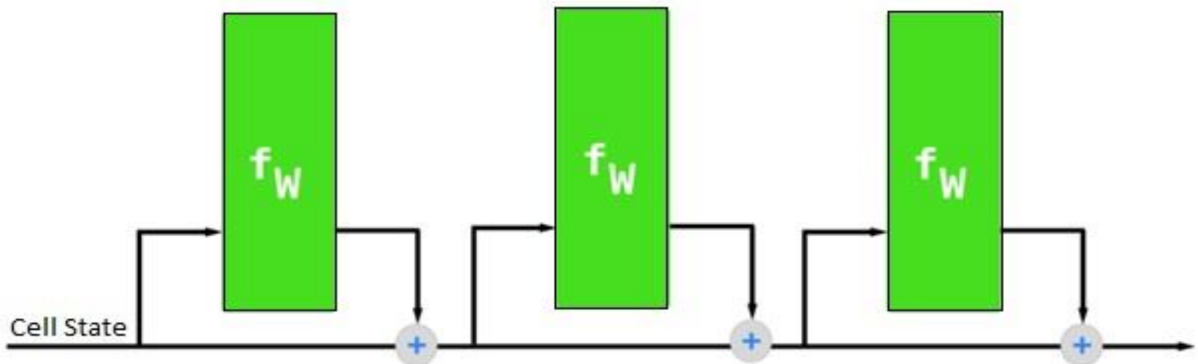


Fig 3.2 RNN cell state representation

A) Previous cell state that is the information that was present in the memory after the previous time step.
B) Previous hidden state that is this is the same as the output of the previous cell.
C) The input at the current time step that is the new information that is being fed in at that moment

3.2 Architecture of LSTM model

A typical LSTM network is comprised of different memory blocks called cells. In Fig. 3.3 There are two states which are being transferred to next cell, the cell state and the hidden state. The memory cells are responsible and capable for remembering content and updation to this memory is done by three important mechanism, that are called as forget gate, input gate and output gate.

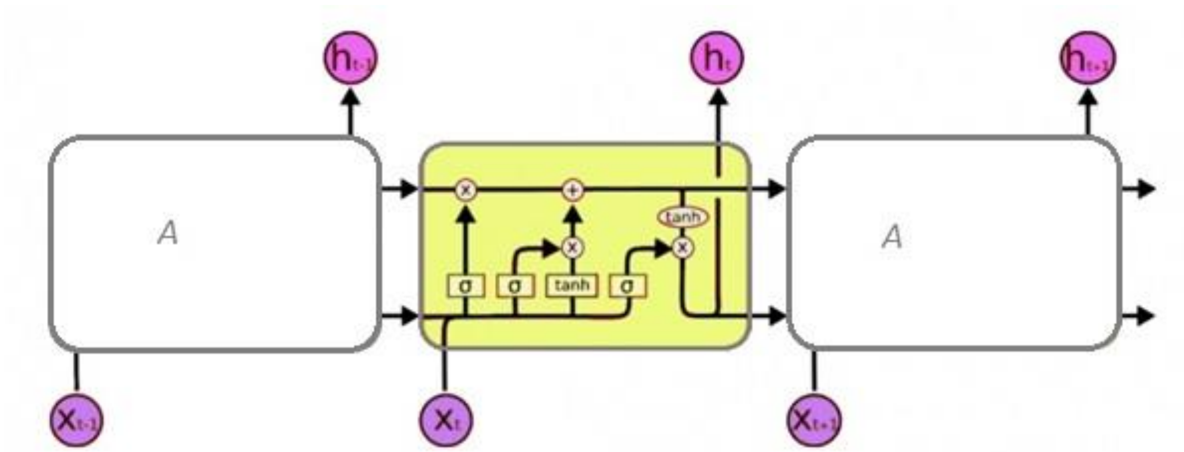


Fig 3.3 Architecture of LSTM model

Forget Gate

A forget gate is capable to remove information from the cell state. The information which is no longer useful or important and not require for the LSTM model is being removed with the help of multiplication of a filter. The forget gate is optimize the overall performance of the LSTM model. For example "John is good person. Alice on the other hand is evil" As soon as the first full stop after "person" is came, the forget gate realizes that there might be some change of context in the upcoming next sentence. As a result of this, the subject of the sentence is forgotten and the place for the subject is vacated. And when we start speaking about "Alice" the forget gate should forget the past conception about the "John". The mechanism of forgetting the stuff is the main responsibility of the forget gate.

Input Gate

The input gate able to add the information to the cell state. The mechanism of addition of information consist of three step process and that are as follows.

Step 1- Regulating what values need to be added to the cell state by involving a sigmoid function.

Step 2- Creating a vector containing all possible values that can be added to the cell state. This is done using the tan h function, which outputs values from -1 to +1.

Step 3- Multiplying the value of the regulatory filter to the created vector and then adding this useful information to the cell state via addition operation.

Output Gate

Creating a vector after applying tanh function to the cell state, thereby scaling the values to the range -1 to +1.

Making a filter using the values of h_t-1 and x_t, such that it can regulate the values that need to be output from the vector created above. This filter again employs a sigmoid function.

Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as a output and also to the hidden state of the next cell.

Chapter 4

Methodology

4.1 Bi-directional LSTM

Bidirectional LSTMs are an extension to normal LSTMs that could beautify the overall performance of the model on sequence classification problems. Where all time steps of the enter series are available, Bi-LSTMs teach or train the LSTMs rather than one LSTMs at the input series. The first at the input series as-is and the opposite on a reversed replica of the input series. By this extra context is introduced to the network and consequences are faster.

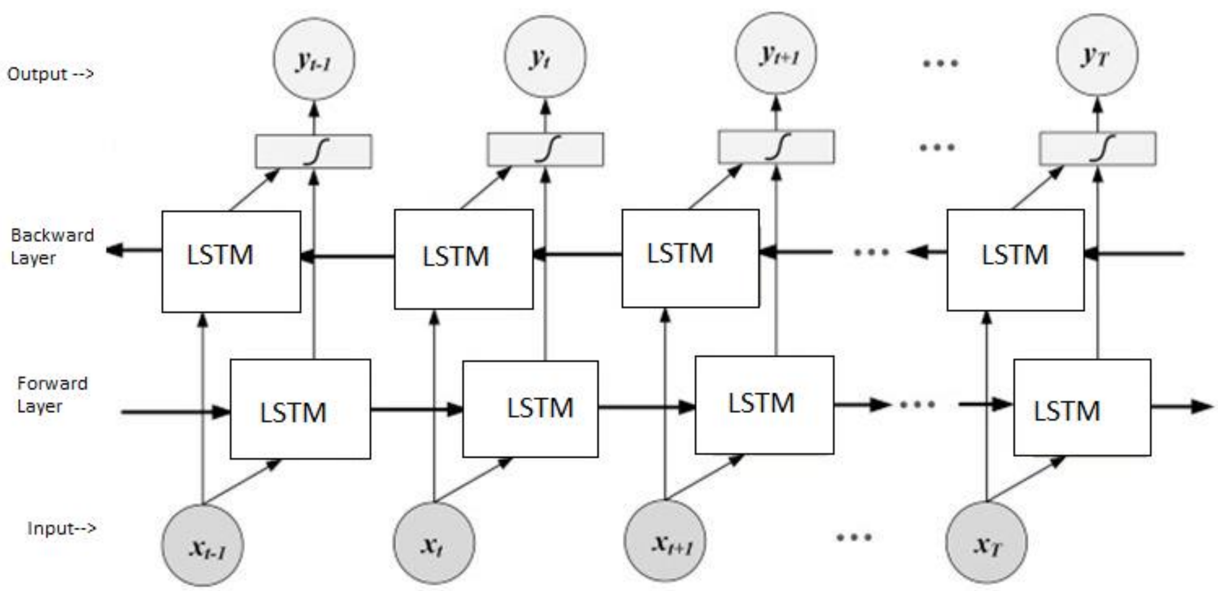
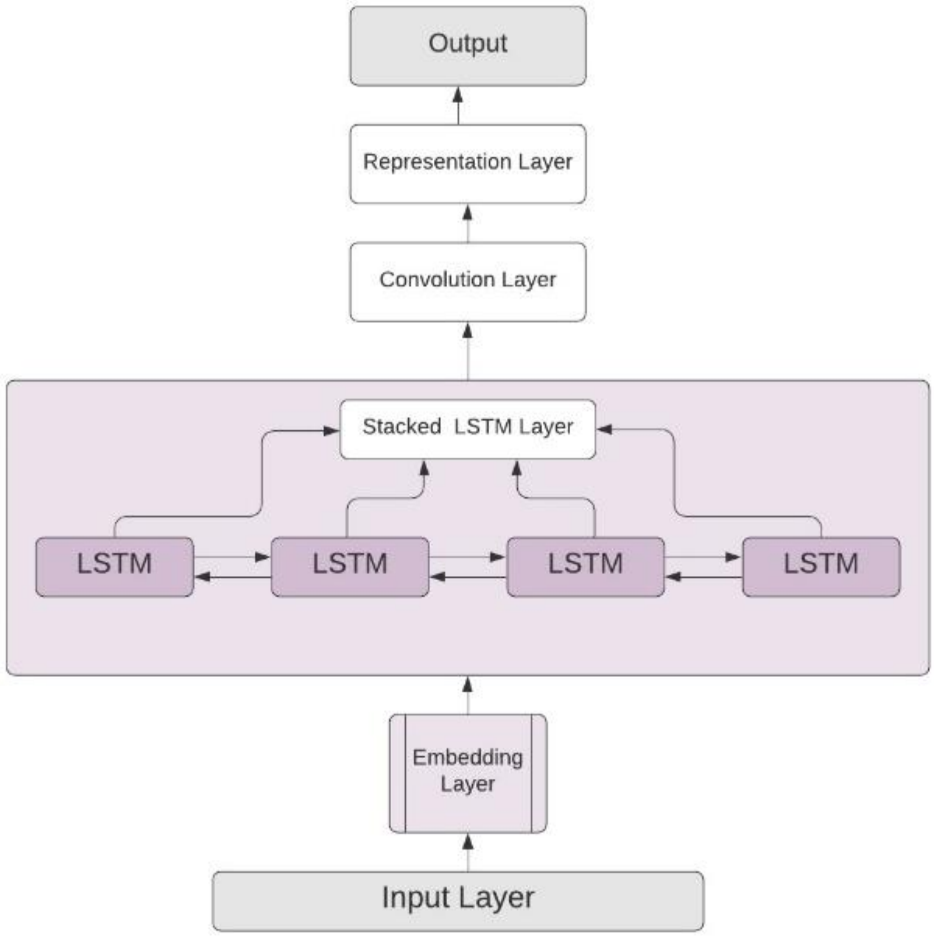


Fig 4.1 Architecture of Bi-Directional LSTM model

Bidirectional Recurrent Neural Networks (RNNs) are based on a simple concept. This entails replicating the network's first recurrent layer, then feeding the input sequence to the first layer as is and a reversed duplicate of the input sequence to the replicated layer. This circumvents the drawbacks of a regular RNN. BRNN (bidirectional recurrent neural network) is a type of recurrent neural network that can be used.

4.2 Proposed Flow Chart

The main focus is on social media as many people especially social activists make use of sarcasm in their tweets or post to express their point of view, and detecting the presence of sarcasm in a particular set of words is not an easy task for humans as well as machine. For example “It is a great feeling to wear face mask and specs together” it is containing sarcasm. expressing negative sentiment about the situation of wearing the face mask and specs together using positive opinion words like ‘‘great feeling’’ this makes sarcasm detection an important task in the field of NLP to improve the overall performance to analyze sentiments effectively.



4.2 Architecture of Proposed Stacked Bi-Directional LSTM model

As part of this study, we proposed a deep learning model for sarcasm detection. The objective and working principle of this work can be elaborated as.

• Input Layer: The dataset is collected from the two news website it contains many extra columns which are not useful while training our model so we removed the extra amount of data from the dataset. After pre-processing of the dataset it is being fed into the input layer. The input layer is nothing but a layer that applies some machine learning algorithm to fetch required data from the dataset. The input layer is connected to the word embedding layer, which builds word embeddings using the GloVe model.

• Embedding Layer: In this layer set of all words from the dataset are being mapped into a vector with a low dimension using the GloVe model. The glove model is responsible to decode every single word in a standard word format that has similar and significant related meaning. For mapping, it used a large word set file that picks a single word and maps it into the respective matched word. After that, it encodes into a real-valued vector which is known as an embedding vector. To map these words to the desired word a look-up table is maintained which facilitates word representation In this proposed work to construct word embedding GloVe is utilized, which produces a word vector table. The gloVe is a word vector methodology that rode the surge of word vectors after a short time. Just to rewrite, word vectors put words to a respectable vector space, where equivalent words bundle together and different words repel. It is a check-based model of addressing words by highlight vectors. This log-bilinear model investigates the relationship of words by telling the occasions they happen. With the goal that the model assistance in planning all the tokenized words in each sentence to their comparing word vector tables. Additional padding is applied to bind together the feature vector matrix.

• BLSTM layer: LSTM is a deep recurrent neural network (RNN) architecture that is used in the deep learning model. Simple LSTM network can be considered as unidirectional LSTM on the other hand BLSTM (bi-directional LSTM) the input runs in two ways and that is from future to past and from past to future so basically, by using these two-way states outcomes the model can successfully collect information from past as well as future. The output of the previous layer that is word-embedding is then fed into the BLSTM layer.

• Stacked LSTM layer: Stacking multiple LSTM layers together makes the model more accurate at the same time more complex, when we work with such a type of deep learning model flow of data should be clear and properly feed into the next layer of our model take care of all such type of scenario.

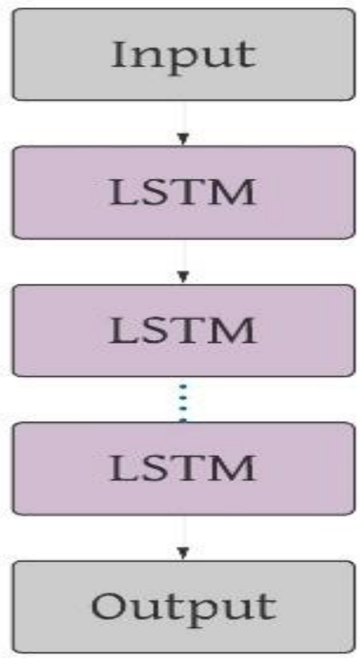


Fig. 4.3 Flow Chart of Stacked based bi-directional LSTM.

• Convolution Layer: The convolution layer play an important role in building any convolution neural network model . A CNN (Convolution neural network) is a subpart of deep neural network which is mostly use to analyse data like images. Finally the output of the stacked based LSTM combined with auxiliary features and this is provided as the input to this convolution layer.

• Representation Layer: It is the last layer of out model which do nothing but to represent the final output based on the input which is passed by different layers , It is a connected layer that make use of linear transformation method using softmax activation function. Which able to calculate the probablity of the particular event on the bases of that out model predict that whether a given input is sarcastic or not.

Chapter 5

Experiment and Results

5.1 Dataset

In the age of information, technology data play an important role, data is nothing but a set of value either in the form of structured or non-structured. The data which comprises meaningful values that are collected from the past experiments and collected in a structured manner, collection of this type of data is known as the dataset. We use part of the dataset to train over the model and part of the dataset to test the efficiency of our proposed model. We are using the dataset which is available on kaggle.com which is the data science community that collects data worldwide and also provides several powerful tools and resources like a different kind of dataset. This dataset contains 28000 text data in the form of the table comprises many columns and for every tuple, the text is categorized into two, sarcastic or not-sarcastic. In the past studies for the sentiment analysis purpose generally, twitter dataset is being used, twitter data are noisy in natures as in tweet many hashtags are used. The dataset is collected from two news websites all the real headlines of current events are collected and then categorized into to sarcastic and not-sarcastic.

5.2 Word Cloud

Data visualization is one of the important tasks before giving input to our model. For data visualization, we are using word cloud, which is a technique that represents text data in the form of a collection of words based on their importance and frequency. The highly significant words are represented in the highlighted manure by using different sizes for a different words. Word cloud is generally used to visualize social media data. In the proposed work the data is categorized into two categories and which are sarcastic and not-sarcastic.

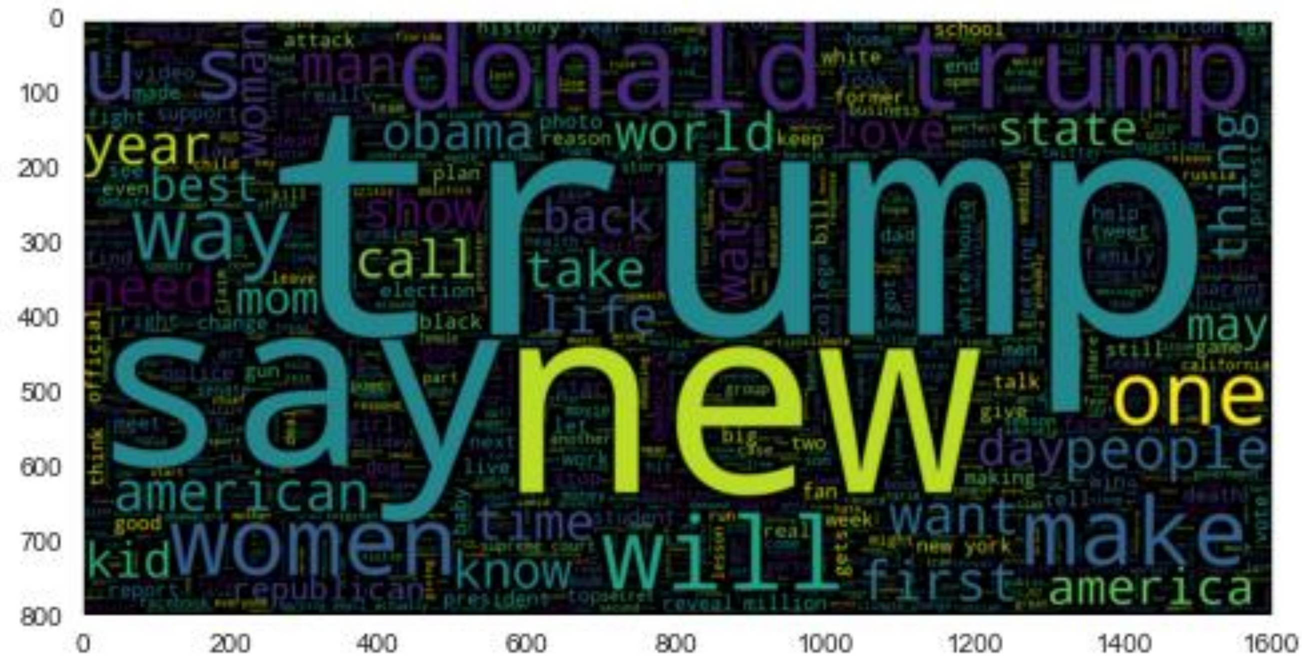


Fig 5.1 Word cloud for not-sarcastic text.

In the proposed work the data is categorized into two categories and which are sarcastic and not-sarcastic, In Fig. 3 the word cloud of not-sarcastic data shows that particular words which are used frequently and have significant importance to detect sarcasm in a given sentence for example in the case of not-sarcastic the word cloud highlighted words like ‘trump’, 'say’, ’new’ and women it shows these words are used the most in not-sarcastic text. Similarly Fig. 4 represented word cloud representation of sarcastic data from the dataset which highlighted words like ‘man’, ‘new’, ‘report’, and nation.

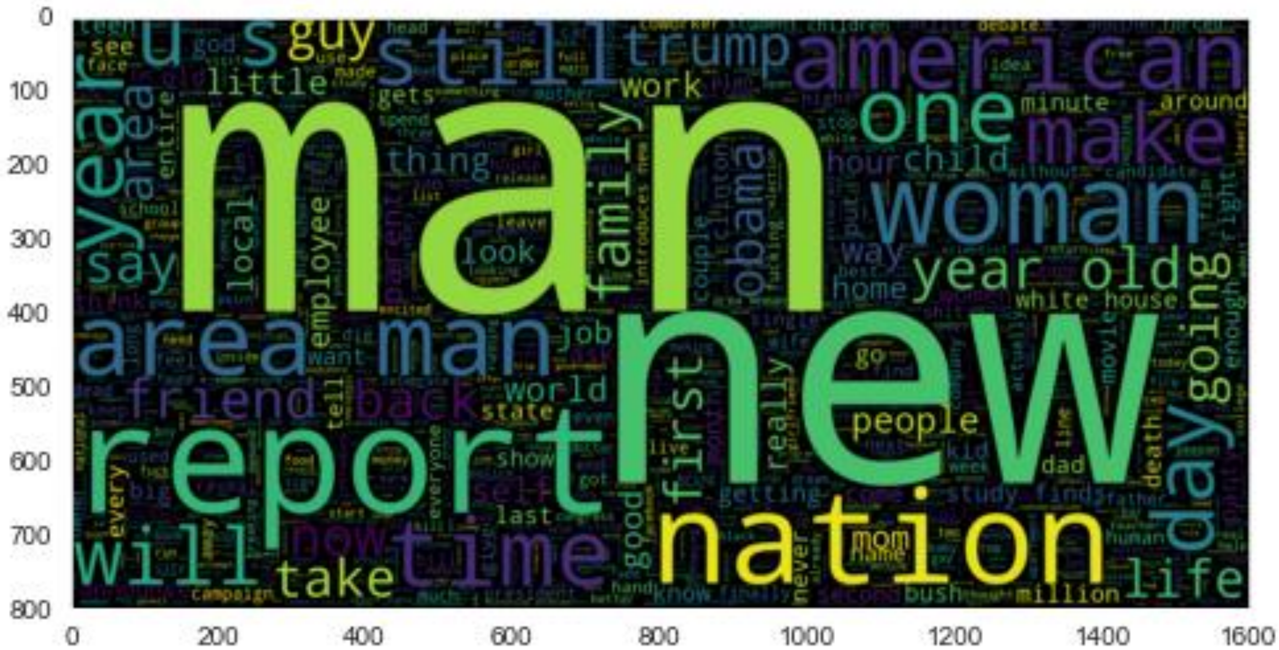


Fig 5.2 Word cloud for sarcastic text.

5.3 Word Embedding

In natural language processing, Word embedding is used for the representation of words for analysis, In the form of a real-valued vector that decodes the meaning of the word such that the words have the same or similar meaning in the vector space. In the proposed work for word embedding, we have used the GloVe model. In this type of word embedding data that show similar meaning are considered as a single phrase. Many words might be left from this real-valued vector so that the GloVe model initializes such type of word using uniform distribution sampling.

5.4 Performance Matrix

5.4.1 Accuracy

It is the unit that is being used to measure the efficiency of any particular model. It shows how accurate our model is based on the error rate. The accuracy of any model is

calculated by this formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Where, TP, TN, FP, and FN are respectively true positive, true negative, false positive, and false negative.

5.4.2 Precision

Precision is calculated by the correctness of the input which is positive in our case that is sarcastic only, not-sarcastic data is not included in finding precision. The precision of any model is calculated by this formula:

$$Precison = \frac{TP}{TP+FP} \qquad (2)$$

5.4.3 Recall

In recall, we compare positive correct that is true positive value with actually positive that is with a truly positive and false negative. The recall is being calculated using the following formula:

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

5.4.4 F-Score

Finally, the F1 score that is nothing but a average between precision and recall but it is a different kind of average, a special type of mean that is called harmonic mean.

$$F1\ score = 2 \,.\, \frac{pricision\,.\,recall}{pricision + recall} \qquad (4)$$

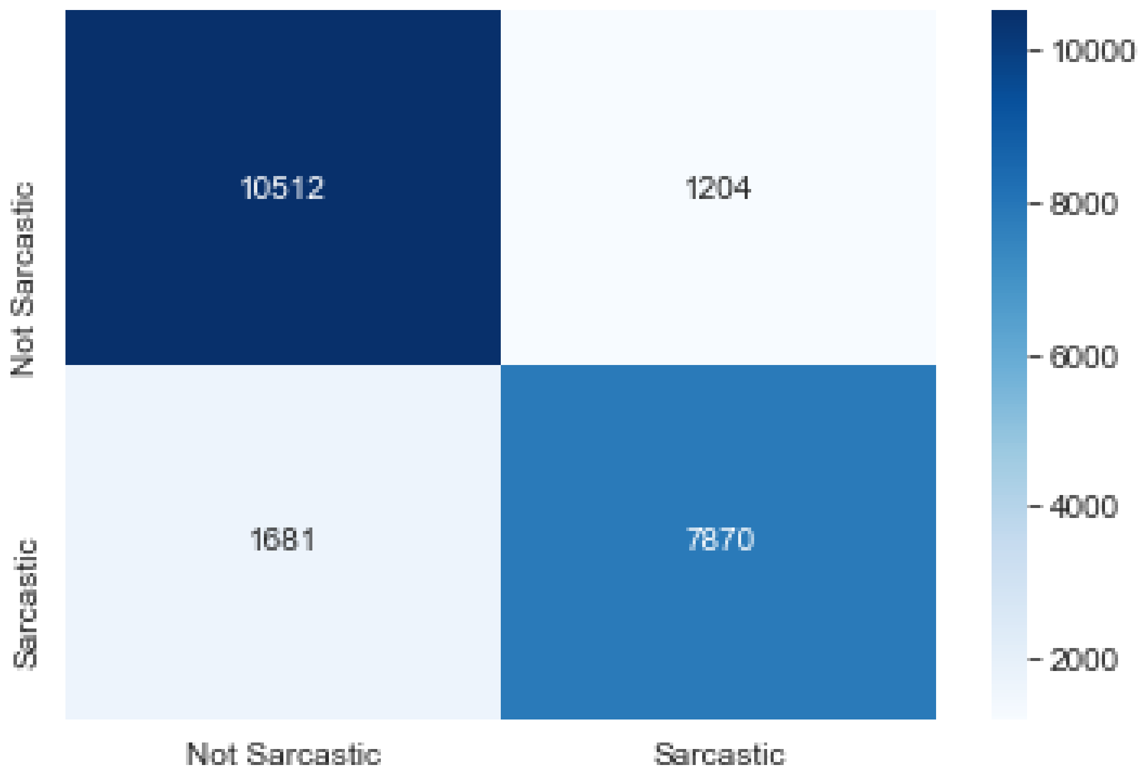


Fig 5.3 Heatmap graph of all parameters TP, TN, FP and FN.

In Fig. 5 Heatmap graph is presented which shows the values of true positive i.e the sarcastic input and predicted sarcastic, true negative that means input which is not-sarcastic and predicted correctly, false-positive implies sentences which are sarcastic but predicted wrong i.e not-sarcastic, false-negative that implies the sentence which is not-sarcastic and predicted wrong i.e sarcastic. The number line shows the amount of data on which model is tested.

TABLE I. Performance matrix obtain from complete tested dataset.

	Performance (in %)
Accuracy	86.47
Precision	82.39
Recall	86.73
F-score	84.50

TABLE II. Performance matrix obtain from complete dataset.

Prediction →	**0 (not-sarcastic)**	**1(sarcastic)**
Accuracy	86.21	86.73
Precision	84.66	80.12
Recall	87.81	85.65
F-score	85.79	83.21

Table I shows the performance matrix which is obtain from complete tested dataset. In our model 80% of data is used for testing. Table II shows the performance parameter on the basis of two class which are ‘sarcastic’ and ‘not-sarcastic’.

The proposed model is also compared with two deep learning model, simple LSTM approach and Bi-LSTM Model. The graph in Fig. 6 gives the comparison of these two deep learning models to the proposed model.

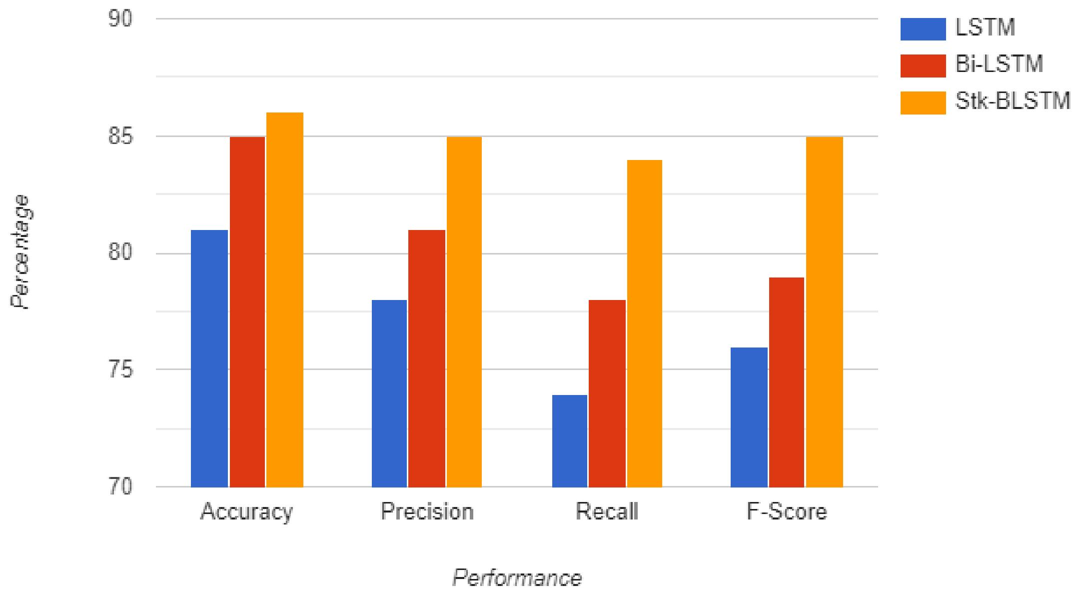


Fig 5.2 Comparison of proposed model with two LSTM based deep learning model.

TABLE III Comparison of proposed model with two LSTM based deep learning model.

Methodology	**LSTM**	**B-DLSTM**	**Stk-BLSTM**
Accuracy	86.21	81.75	86.47
Precision	84.66	78.28	82.39
Recall	87.81	74.91	86.73
F-score	85.79	76.45	84.50

The results demonstrate that the proposed Stacked bi-directional LSTN model achieves the highest accuracy of 86.62% as compared to the other variants. The order of models from highest to lowest accuracy is, Stacked Bi-directional LSTM > Bi-directional LSTM > Basic LSTM model. The proposed Stacked Bi-directional LSTM model also achieves the highest recall, precision and F-Score.

Chapter 6

6.1 Conclusion

In this proposed work, a model is introduced which is capable to detects sarcasm and provide the information that weather a giver set of word or a sentence is sarcastic or not-sarcastic. In this proposed model we used Stacked bi-directional LSTM which is using several layers of Bi-Directional LSTM to get a better result and also used the features like word embedding and convolution layer. The main aim of this proposed work is to apply more information in the present Bi-LSTM models and we successfully achieved improvement in the form of performance matrix by stacking the layer or Bi-LSTM.

Basically our goal in this work is to find the public opinion and perform the opinion mining. Generally what happens, through post on social media people express their thoughts, feelings etc. but we could not able to find the people thoughts and feelings. So by performing sentiment analysis on those sarcastic posts finally we can conclude how many person are in favor of this mission and how many person are against of this mission.

6.2 Future Work

In future work, we can work on different type of dataset like multimedia and video based data after adding features like video summarization using deep learning approach. Future scope also includes that we can make web application for our work. In addition to this we can improve our classifier system such that it could deals with sentences that conveys multiple meaning. Furthermore, we can add more classification categories so that we could get better results. We can also design system such that it can detect the images in dataset with the help of image processing and also able to detect the sarcasm in multimedia and video based dataset.

Appendix A

Appendices

A.1 Publication (Communicated)

A.1.1 Sarcasm Detection Using Stacked Bi-Directional LSTM Model

Abstract - Sarcasm is one of the sentiments which is being used to communicate a negative opinion utilizing positive words. The world is full of social media and many kinds of the web-based portal and this media stores a huge amount of textual data which contains sentiments, and sarcasm is one of the sentiments which is being used nowadays in many of this platform, using sarcasm someone can communicate their negative words in a positive way which is we can call a sarcastic way of communication. In opinion mining, the field of natural language processing detection of sarcasm from a given data is an important task. It is a binary classification task for which model proposed a system which classifies whether a given set of word is sarcastic or not-sarcastic. In this research work, we proposed the work based on the Stacked Bi-Directional Long Short-Term Memory (Stk-BLSTM) network which enhances the overall result in terms of performance matrix.

Keywords - Sarcasm detection, Natural Language Processing, Stacked LSTM, Machine learning, Social data.

Bibliography

[1] Yufeng Diao, Hongfei Lin, Liang Yang, Xiaochao Fan, Yonghe Chu, Kan Xu, And Di Wu. “A Multi-Dimension Question Answering Network for Sarcasm Detection” IEEE Acess, 2021

[2] https://backlinko.com/social-media-users

[3] A. Ghosh and T. Veale, ‘‘Fracking sarcasm using neural network,’’ in Proc. 7th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal, 2016

[4] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, ‘‘Sarcasm as contrast between a positive sentiment and negative situation,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., 2013, pp. 704–714.

[5] D. I. H. Farias, ‘‘Irony and sarcasm detection in Twitter: The role of affective content,’’ ACM Trans. Internet Technol., vol. 16, no. 3, p. 19, 2016.

[6] A. Kumar and A. Jaiswal, ‘‘Empirical study of Twitter and Tumblr for sentiment analysis using soft computing techniques,’’ in Proc. World Congr. Eng. Comput. Sci., 2017, pp. 1–5.

[7] C. W. Tseng, J. J. Chou, and Y. C. Tsai, ‘‘Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members,’’ IEEE Access, vol. 6, pp. 72870–72879, 2018.

[8] D. Zhang, L. Tian, M. Hong, F. Han, Y. Ren, and Y. Chen, ‘‘Combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification,’’ IEEE Access, vol. 6, pp. 73750–73759, 2018.

[9] X. Fu, J. Yang, J. Li, M. Fang, and H. Wang, ‘‘Lexicon-enhanced LSTM with attention for general sentiment analysis,’’ IEEE Access, vol. 6, pp. 71884–71891, 2018.

[10] M. Bouazizi and T. O. Ohtsuki, ‘‘A pattern-based approach for sarcasm detection on Twitter,’’ IEEE Access, vol. 4, pp. 5477–5488, 2016.

[11] Le Hoang son , Akshi Kumar, Saurabh Raj Sangwan, Anshika Arora, Anand Nayyar , and Mohamed Abdel-basset, “Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model With Convolution Network” IEEE Access 2019.

[12] Avinash Kumar , Vishnu Teja Narapareddy , Veerubhotla Aditya Srikanth , Aruna Malapati , and Lalita Bhanu Murthy Neti, “Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM” IEEE Access 2019.

[13] Yufeng Diao, Hongfei Lin, Liang Yang , Xiaochao Fan, Yonghe Chu, Kan Xu, and Di Wu, “A multi-dimension question answering network for sarcasm detection” IEEE Access 2020.

[14] T. Veale and Y. Hao, ‘‘Detecting ironic intent in creative comparisons,’’ in Proc. ECAI, vol. 215, 2010, pp. 765–770.

[15] D. Maynard and M. A. Greenwood, ‘‘Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis,’’ in Proc. LREC, 2014.

[16] S. K. Bharti, K. S. Babu, and S. K. Jena, ‘‘Parsing-based sarcasm sentiment recognition in Twitter data,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, 2015, pp. 1373–1380

[17] Gonzalez-Ibanez R, Wacholder N. Identifying sarcasm in twitter: a closer look. In Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies: Short s-Volume 2. 2011;p 140–142.

[18] Sivaprakasam L. A study on sarcasm detection algorithms. Int J Eng Technol Sci Res. 2017;76–84.

[19] Chandra N. Sentiment analysis and its challenges. In: World congress on computing and communication technologies (WCCCT). 2017. p. 140–142.