

# Sentiment Analysis Using Deep Learning Approach

Submitted By

Sanjida Akter Ishita	170204089
Md Ashfaqul Azam Chowdhury	170204097
Fariha Zaman	170204098

Submitted To

**H M Zabir Haque**, Assistant Professor  
**Nibir Chandra Mandal**, Lecturer  
Department of Computer Science and Engineering

Project Report

Course ID: CSE 4238

Course Name: Soft Computing Lab

Semester: Spring 2021



**Department of Computer Science and Engineering**  
**Ahsanullah University of Science and Technology**

Dhaka, Bangladesh

15 March 2022

# ABSTRACT

Deep Learning algorithms are able to identify and learn the patterns from both unstructured and unlabeled data without human intervention. Deep Learning techniques learn through multiple layers of representation and generate state of the art predictive results. In the past years, Deep Learning techniques have been very successful in performing the sentiment analysis. Mature deep learning neural network has completely changed the field of natural language processing (NLP). Due to the enormous amount of data and opinions being produced, shared and transferred everyday across the Internet and other media, sentiment analysis has become one of the most active research fields in natural language processing. This paper introduces three deep learning networks applied in IMDB movie reviews sentiment analysis. The name our dataset is IMDB Dataset of 50K Movie Reviews. Dataset was divided to 50% positive reviews and 50% negative reviews. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) neural networks are two main types, which are widely used in NLP tasks, while Convolutional Neural Networks (CNN) is often used in image recognition. The results have shown that, CNN network model can achieve good classification effect when applied to sentiment analysis of movie reviews. CNN have reported the accuracy of 88.14666867256165%, while RNN and LSTM have reported accuracy of 85.0600004196167% and 88.53% respectively.

# Contents

<b>ABSTRACT</b>	<b>i</b>
<b>List of Figures</b>	<b>iii</b>
<b>List of Tables</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Definition . . . . .	1
1.2 Summarized idea . . . . .	1
<b>2 Motivation</b>	<b>2</b>
2.1 Why our project important . . . . .	2
2.2 Aspects of society . . . . .	2
<b>3 Methodology</b>	<b>4</b>
3.1 Model . . . . .	4
3.1.1 CNN . . . . .	4
3.1.2 RNN . . . . .	5
3.1.3 LSTM . . . . .	5
<b>4 Experiments</b>	<b>7</b>
4.1 Dataset Collection . . . . .	7
4.2 Statistic of dataset . . . . .	7
4.3 Description of data . . . . .	8
4.4 Performance Metrics . . . . .	8
4.5 Results/Evaluation . . . . .	10
4.5.1 Performance of model . . . . .	10
4.5.2 Comparison . . . . .	10
4.5.3 Result . . . . .	10
4.5.4 Discussion . . . . .	10
<b>References</b>	<b>11</b>

# List of Figures

3.1	Convolution Works [1] [2]	4
3.2	RNN [3]	5
3.3	Long short term memory [4]	6
4.1	Confusion matrices of CNN	8
4.2	Confusion matrices of RNN	9
4.3	Confusion matrices of LSTM	9

# List of Tables

4.1 Training result comparison . . . . .	10
--	----

# Chapter 1

## Introduction

### 1.1 Definition

In our project, sentiment analysis is the process of retrieving information about a consumer's perception of a movie..By definition, sentiment analysis or opinion mining is the use of text analysis, computational linguistics, or Natural Language processing (NLP) in order to get semantic quantification of the studied information. Sentiment analysis or opinion mining is the analysis of people's opinions, emotions, and attitudes from written expressions.

### 1.2 Summarized idea

In this project uses the deep learning technology to analyze the sentiment of the movie reviews in the IMDB dataset, and divide them into negative and positive categories. It is necessary to classify movie reviews. For the user, it is a recommendation tool for movie selection, which helps the user make a choice. For film companies, this information can be used for marketing decisions and finding customers. This paper is introducing three neural network models (CNN, RNN, LSTM).

# Chapter 2

## Motivation

Sentiment analysis aims to indicate the opinion of a specific text (e.g. a tweet, or a product review). These indications are used accordingly by decision makers in planning and taking appropriate actions such as marketing decisions, customers hunting or business expansion in specific geographic region.

### 2.1 Why our project important

Sentiment analysis is particularly important when anyone's business is making big moves. For example, product launches, pricing changes and another big announcement could see a significant shift in anyone's brand sentiment. Keeping an eye on these numbers on a regular basis can provide you with peace of mind or spot a potential crisis before it occurs.

The ability to get honest feedback so quickly is something we take for granted with social media, and yet it's one of the biggest attributes of sentiment analysis. One of the best uses of this is prior to launching a new product or campaign – to be sure audience even wants what anyone is offerin

Here, our project is important because it classifies movie reviews from the IMDB dataset into positive and negative categories.

### 2.2 Aspects of society

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. Social media monitoring tools

like Brandwatch Analytics make that process quicker and easier than ever before, thanks to real-time monitoring capabilities.

In our project here classifies movie reviews from the IMDB dataset into positive and negative categories. It can also be an essential part of your market research and customer service approach. Not only can you see what people think of movies, It can be seen what they think about anyone's competitors too. The overall audience experience can be revealed quickly with sentiment analysis.



# Chapter 3

## Methodology

### 3.1 Model

#### 3.1.1 CNN

Convolution Neural Network is a type of feed forward network, which consists of two or more layers deep within and then connected with a fully connected layers like a multilayer neural network. In the perspective of sentiment analysis, CNN works on the process in which each word is given a weight in the hidden layer. Further each word is being checked for the exact match and the process continues in a repeated manner. CNN also works based on the logic of sliding window. For an instance, if an image is given each, filters are decided and passed through the image as a sliding window. This gives the corresponding value of the image and is stored as a matrix. Thus for the entire image, a matrix will be calculated. In the case of text classification, every word will be given as an input and finally represented in a matrix format as shown in Figure 1. Feature detection is done by the convolution layers.

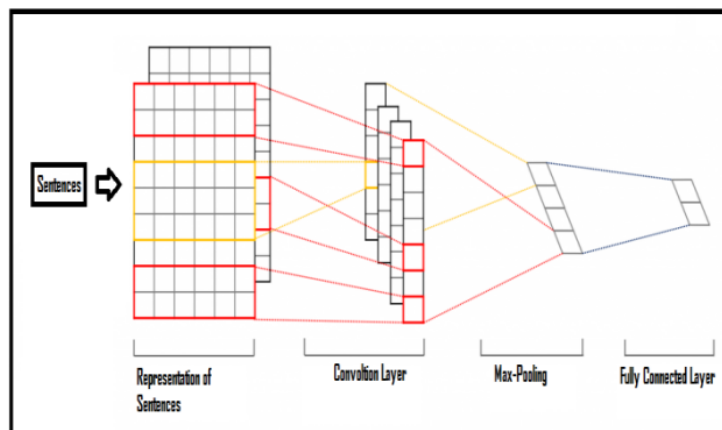


Figure 3.1: Convolution Works [1] [2]

### 3.1.2 RNN

RNN is one of the deep learning approaches which are used for sentiment analysis. It produces the output on the basis of previous computation by using sequential information. Previously, traditional neural network uses independent inputs which are unfit for some task in Natural Language Processing. For an example: word prediction in a given sentence. RNN is efficient model for sentiment analysis. RNN uses memory cell that capable to capture information about long sequences.

In RNNs, if a sequence is long enough, they have a hard time carrying information from earlier computations to later ones, which may leave out important information from the beginning. Therefore, these causes the need of Long Short Term Memory (LSTM) which is a special kind of RNN's, capable of learning long-term dependencies. Some common LSTM applications include time series prediction, speech recognition, video activity recognition, and many more. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

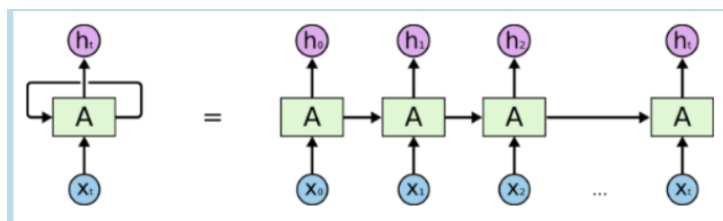


Figure 3.2: RNN [3]

### 3.1.3 LSTM

LSTM (Long Short Term Memory) is the part of RNN which is used to learn long-range dependencies for text sequences. LSTM contains memory blocks which also known as gates to control the text flow. The memory blocks contain three gates named as input gate; forget gate and output gate to control the flow of information.

In RNNs, if a sequence is long enough, they have a hard time carrying information from earlier computations to later ones, which may leave out important information from the beginning. Therefore, these causes the need of Long Short Term Memory (LSTM) which is a special kind of RNN's, capable of learning long-term dependencies. Some common LSTM applications include time series prediction, speech recognition, video activity recognition, and many more. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three

gates regulate the flow of information into and out of the cell.

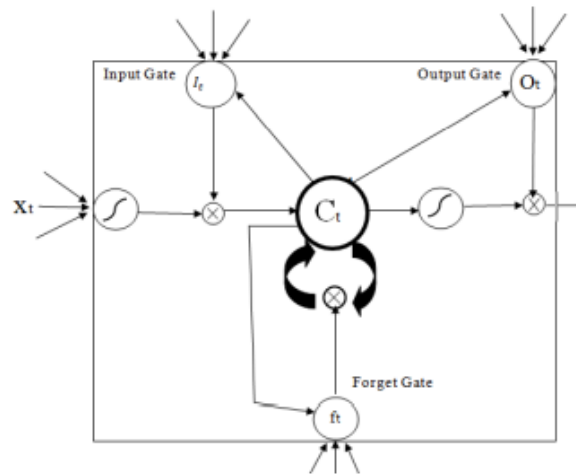


Figure 3.3: Long short term memory [4]

# Chapter 4

## Experiments

### 4.1 Dataset Collection

The dataset which is used for this project is provided by Lakshmipathi N. This dataset is available on Kaggle and it is open for everyone. The dataset is processed before analysis. This dataset is most reviewed dataset for Text Analytics. The dataset link is here - <https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews/version/1>

### 4.2 Statistic of dataset

The dataset is covered by 50000 data. The dataset is divided into training set and test set accordingly 25,000 and 25,000 data. The dataset is about text analysis. In analysis part, the sentiment from sentiment column is dependent on review column. The sentiment part is decided on basis of the review texts. The movie review column has some focused words which are counted as negative words and also some words which are counted as positive words. On basis of this, The analysis of sentiment is decided. Using deep learning algorithm the sentiment column is decided what it will be negative or positive.

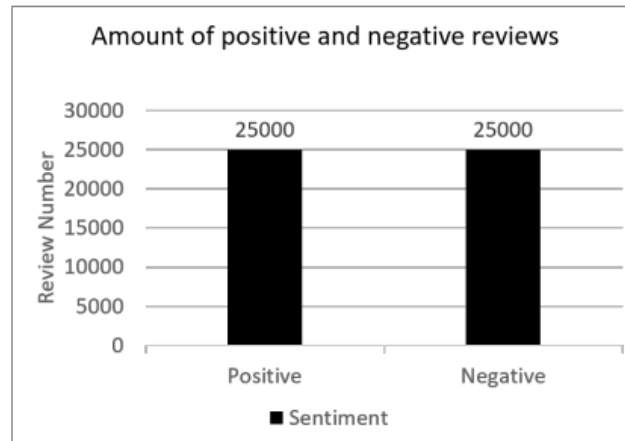


Fig. 7. Statistics of reviews

### 4.3 Description of data

This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. The sentiment and review column are the two columns of the dataset. In review the sentences are countered for the sentiment column. The sentiment part is the target part in this dataset. If a sentence has more positive words in review part then the sentiment will be positive ultimately or it will be negative. The models are applied on this dataset to know the sentiment eventually.

### 4.4 Performance Metrics

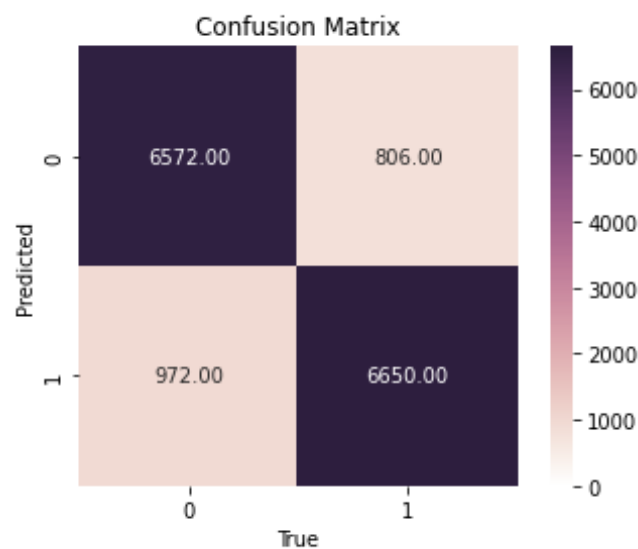


Figure 4.1: Confusion matrices of CNN

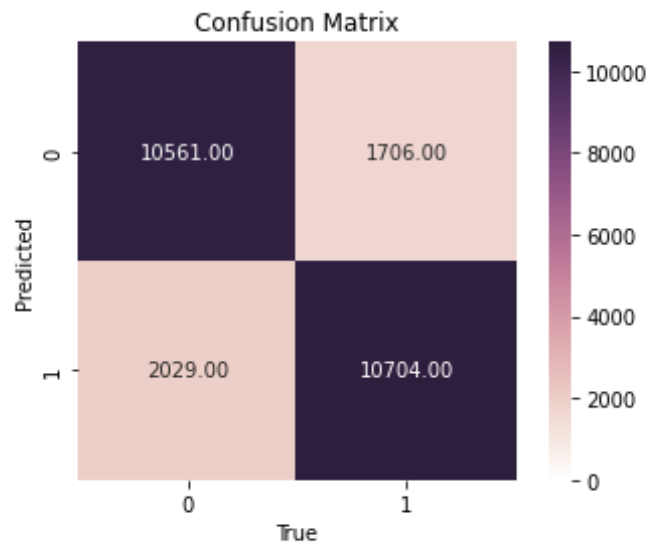


Figure 4.2: Confusion matrices of RNN

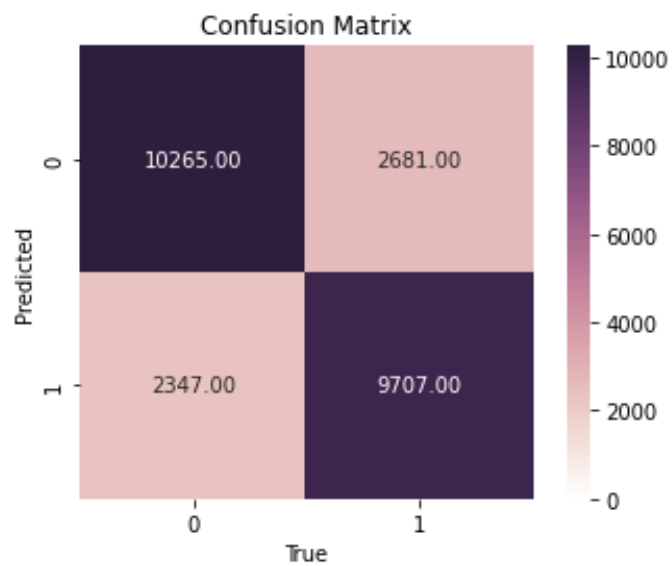


Figure 4.3: Confusion matrices of LSTM

## 4.5 Results/Evaluation

### 4.5.1 Performance of model

### 4.5.2 Comparison

The comparison between the paper result and our result is shown in Table 4.1.

Models	Paper Accuracy	Our Accuracy
CNN	88.22%	88.14666867256165%
RNN	68.64%	85.0600004196167%
LSTM	85.32%	88.53%

Table 4.1: Training result comparison

### 4.5.3 Result

From the result comparison we can see that each one of our model performed slightly better than the paper performances and CNN model performed quite like as the paper performance.

### 4.5.4 Discussion

In this project, we 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. Through experiment, We found that the LSTM still has good performance for processing a sequence of data, achieving an accuracy of 88.53%.

For future work, we hope to create our own integration through the method of superposition model, set up better models in the experiment, and improve the effect of movie review or other emotional analysis. We also know that the data preprocessing greatly affects the model recognition and feature extraction. Therefore, it is the focus of our future work to find better data preprocessing methods.

## References

- [1] B. S. Lakshmi, P. S. Raj, R. R. Vikram, *et al.*, “Sentiment analysis using deep learning technique cnn with kmeans,” *International journal of pure and applied mathematics*, vol. 114, no. 11, pp. 47–57, 2017.
- [2] P. Cen, K. Zhang, and D. Zheng, “Sentiment analysis using deep learning approach,” *J. Artif. Intell*, vol. 2, no. 1, pp. 17–27, 2020.
- [3] R. C. Staudemeyer and E. R. Morris, “Understanding lstm—a tutorial into long short-term memory recurrent neural networks,” *arXiv preprint arXiv:1909.09586*, 2019.
- [4] A. Patel and A. K. Tiwari, “Sentiment analysis by using recurrent neural network,” in *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, 2019.



Generated using Undergraduate Thesis L<sup>A</sup>T<sub>E</sub>X Template, Version 1.4. Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.

This project report was generated on Tuesday 15<sup>th</sup> March, 2022 at 5:31pm.