

# Analysis of hyperparameters in Sentiment Analysis of Movie Reviews using Bi-LSTM

Amankumar Singh<sup>1,\*</sup>, Riya Thapliyal<sup>1,\*\*</sup>, Ritika Vanave<sup>1,\*\*\*</sup>, Rajashree Shedge<sup>1,\*\*\*\*</sup>, and Snehal Mumbaikar<sup>1</sup>

<sup>1</sup>Computer Engineering Department, Ramrao Adik Institute of Technology, Navi Mumbai, India.

**Abstract.** Movie reviews are an important factor in determining a film's success because instead of depending solely on the number of views as a parameter for the success of the movie, movie reviews are used to acquire additional insights into the movie. Existing systems use LSTM for sentiment analysis but there is no study available how various hyperparameters affect the performance of the model. Bi-LSTM along with dropout layers provide good accuracy in sentiment analysis. The suggested method outperforms CNN and Natural Language Toolkit in terms of accuracy. The proposed model is tested using different hyper parameters including dropout rate, number of Bi-LSTM layers and Bi-LSTM nodes. 64 LSTM nodes, 2 Bi-directional Layers, and a 0.2 Dropout rate should be used for optimal accuracy. Effect of different text vectorization algorithms and activation functions was also studied. The combination of Tf-idf text vectorization and the ReLU activation function yields the best results.

**Keywords:** Sentiment Analysis, Text-vectorization, Bi-LSTM.

## 1 Introduction

Sentiment Analysis on movie reviews can help readers make more informed decisions by providing a more comprehensive description of the film[1]. Sentiment analysis is used to get the performance of the movie based on user reviews. Any sequence of words that is usually associated with a specific opinion is referred to as a sentiment. Sentiment analysis not only focuses on polarity but also on feelings, emotions, and intentions. This type of analysis can be extremely beneficial in marketing, customer feedback, and customer service. In this scenario, using machine learning to learn and predict whether a movie review is positive or negative, the analysis is the act of looking at data and generating judgments. Two main strategies for sentiment analysis are rule-based and machine learning-based. The machine learning model identifies sentiment based on words and their order, whereas the rule-based technique employs sentiment scores. News and social media datasets can also be used for sentiment analysis.

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network that may learn order dependence in sequence prediction challenges[8]. LSTM preserves historical information by establishing a memory unit which can substantially reduce the long-distance reliance problem of Recurrent Neural Networks. The LSTM is made up of three sections, each of which serves a distinct purpose. The first portion determines whether the previous timestamp's information should be remembered or is irrelevant and can be ignored. The cell attempts to

learn new information from the input in the second section. Finally, the cell passes updated information from the current timestamp to the next timestamp in the third component. The gates are the three components of an LSTM cell[8]. The Forget gate is the first part, the Input gate is the second, and the Output gate is the final part. Future information is ignored by LSTM. BiLSTM[3] adds another LSTM layer to the mix, reversing the flow of data. At a nutshell, it means that in the additional LSTM layer, the input sequence flows backward. The outputs from both LSTM layers are then combined in a variety of methods, including average, sum, multiplication, and concatenation. BiLSTM can provide a more relevant output by merging LSTM layers from both directions, because every component of an input sequence carries information from both the past and the present[3].

Section II of the paper gives a brief description about the literature survey done on the existing systems and comparative analysis of the survey. The problem definition and the need for such a system is covered in section III. Section IV discusses the proposed system and architecture of the system. The result of the system is covered in Section V. Section VI draws the conclusion of the paper.

## 2 Related Work

To train machines to perform sentiment analysis, a variety of techniques and complicated algorithms are used. Each one has their benefits and drawbacks. However, when used in tandem, they can provide extraordinary outcomes. In recent years, there has been a lot of research on sentiment analysis of text based on movie reviews. For sentiment analysis several methodologies were used.

\*e-mail: aks001235@gmail.com

\*\*e-mail: riyathapliyal2018@gmail.com

\*\*\*e-mail: ritikavanave@gmail.com

\*\*\*\*e-mail: rajashree.shedge@rait.ac.in

An ML-based model is proposed by Saeed Mian Qaisar[1], in which the imported dataset is first partitioned, then vectorized, and then submitted to the LSTM model. Three layers are used here: a 50-node first layer, a 101-node second LSTM layer, and a final layer with outputs. This paper uses a 50k movie review dataset from IMDb for its research. To identify IMDb movie reviews, Aswathi Sajeevan and Lakshmi K S [2] utilize two separate deep learning models: the first is a hybrid LSTM-CNN, while the second is a hybrid CNN-LSTM. This study uses the IMDb movie reviews dataset, which includes 1000 positive and 1000 negative reviews. Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu[3] use the TensorFlow library, as well as Dropout and TF-IDF, to develop a BiLSTM model. Comment text vectors are created from the textual data. ReLU is the activation function employed. The paper is based on a crawl of 15000 hotel comment texts from the Ctrip website.

Anwar Ur Rehman, Ahmad Kamran Malik, Basit Raza, and Waqar Ali offer a model with recurrent and convolutional neural networks in their paper[4]. It starts with a corpus and then preprocesses it. Sentiment Analysis was then performed using a word embedding layer in conjunction with CNN and LSTM models. The categorization layer is then applied at the end. The IMDb movie reviews dataset and the Amazon movie reviews dataset were used. The Nave Bayes algorithm, which is a supervised learning approach as well as a statistical method for classification, is used in this study by G. P Saradhi Varma, A.Govardhan, and I.Hemalatha[5]. It also employs the maximum entropy algorithm, which is a machine learning technique for classification and prediction that may be used in a variety of situations.

The text classification approach is utilized in this study by Madhav Singh Solanki[6], which is a supervised learning approach to recognise a specific type of writing, such as a blog, book, web page, news item, or tweet. Another rule based model is used that employs lexicon to recognise certain terms in the text. Rule-based techniques typically define a set of rules in a scripting language that denotes subjectivity, polarity, or a point of view. Machine learning techniques, POS tagging, dependency parsing, and other techniques are utilized in this article by Amalan Chakrabarti[7] to identify the aspects and user opinions linked to the aspect. The input layer, which consists of word embedding characteristics for each word in the phrase, and two convolution layers make up a seven-layer deep CNN architecture.

**Table 2.1.** Comparative Analysis of the existing systems

Techniques	Advantages	Limitations
Sentiment Analysis of IMDb Movie Reviews Using Long Short Term Memory[1]	<ol style="list-style-type: none"> <li>1. The model is very simple to build and understand.</li> <li>2. The accuracy achieved in this paper is 89.9</li> </ol>	<ol style="list-style-type: none"> <li>1. A better precision can be achieved by further cleaning data and employing the ensemble classification approaches.</li> </ol>
An enhanced approach for movie review analysis using deep learning techniques[2]	<ol style="list-style-type: none"> <li>1. The framework utilizes this advantage of both CNN and LSTM.</li> <li>2. High performance.</li> </ol>	<ol style="list-style-type: none"> <li>1. Accuracy of CNN-LSTM is low at 51.75%.</li> </ol>
Sentiment Analysis of Comment Text Based on BiLSTM[3]	<ol style="list-style-type: none"> <li>1. F1 score - 92</li> <li>2. Both forward and reverse sequences are considered.</li> </ol>	<ol style="list-style-type: none"> <li>1. The sentiment analysis using BiLSTM takes a long time for training</li> </ol>
A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis[4]	<ol style="list-style-type: none"> <li>1. Local features extraction.</li> <li>2. Capture long-term dependencies</li> <li>3. Improved the execution time.</li> </ol>	
Sentiment Analysis Using Machine Learning Algorithms[5]	<ol style="list-style-type: none"> <li>1. Machine learning algorithms can achieve high accuracy for classifying sentiment when using this method.</li> <li>2. A leading advantage of maximum entropy (ME) models is their flexibility.</li> </ol>	<ol style="list-style-type: none"> <li>1. ME models requires considerable computational resources.</li> <li>2. Estimation of such models also leads to sensitivity to round-off errors.</li> </ol>

Sentiment Analysis of Text using Rule based and Natural Language Toolkit[6]	<ol style="list-style-type: none"> <li>1. Sentiment Analysis of Text using Rule Based algorithm is fast and has high accuracy.</li> <li>2. To improve analytical speed and reliability, the map reduction environment can be used.</li> </ol>	<ol style="list-style-type: none"> <li>1. One of the limitations of using lexicon resorting is that the way people express their emotions varies enormously.</li> </ol>
A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis[7]	<ol style="list-style-type: none"> <li>1. In comparison with the existing methods, the proposed technique (CNN and Rule-Based) brings better classification accuracy for both the positive and negative classes.</li> </ol>	<ol style="list-style-type: none"> <li>1. The model developed using both the rule based and Machine Learning method is very complex and needs to be tweaked on various aspects for best performance.</li> </ol>

For sentiment analysis, the majority of studies have employed either traditional deep learning algorithms like LSTM and CNN. Models employing CNN or machine learning[6] have a long training time and require large computing power.

### 3 Proposed System

Gauging the performance of the movie based solely on the views of the movie can be mis-leading and understanding the performance of the movie based on user comments will be more realistic. Therefore, a system is proposed that could perform sentiment analysis on the comments of the users and provide users with the rating of the movie based on the comments. The problem is best framed as single-label classification, which generates a numerical measure of positivity and negativity. The comment given by the user will be mapped to a scale of 1 to 5, 1 being the most negative and 5 the most positive. Real-life scenarios like emojis, slang language and spelling mistakes will be taken into consideration while doing the classification. The use of LSTM model to solve sentiment analysis is proposed.

#### A. System Architecture

The proposed system architecture has 5 phases beginning from text vectorization, then model building and evaluation followed by input sanitization. Based on the input, the model calculates a sentiment score.

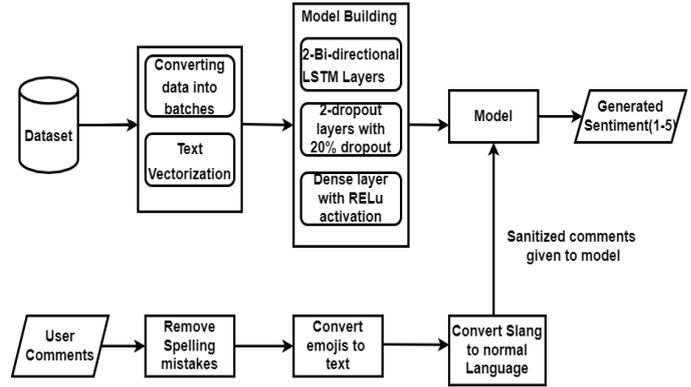


Figure 3.1. Proposed Framework

#### B. Proposed Methodology

1. Stage 1(Data preparation) : Convert data into batches for better training and perform text vectorization on data.
2. Stage 2(Model Building): 2 Bi-directional LSTM layers along with two dropout layers with a 20% dropout rate are used. At the output layer, use the Dense layer with the ReLu Activation function.
3. Stage 3(Model Evaluation): Model is evaluated on Binary cross entropy loss and Adam Optimizer. Metric used is accuracy.
4. Stage 4(Input Sanitization): Spelling errors are removed and emojis are converted into text.
5. Stage 5(Feeding data to model and Output generation): The entire sanitized input is then sent to the model and it generates the output.

#### C. Bi-LSTM

The proposed system uses Bi-LSTM. LSTM networks are designed to overcome the long-term dependency problem faced by recurrent neural networks. LSTMs can process entire sequences of without having to handle each point in the sequence separately, instead, preserving useful information about previous data in the sequence to help with the processing of new data points. LSTMs employ a number of 'gates', which regulate how data in a sequence enters, is stored in, and exits the network. A typical LSTM has three gates: a forget gate, an input gate, and an output gate. These gates are each their own neural network and can be thought of as filters. A Bi-LSTM is a sequence processing model made up of two LSTMs, one of which takes the input in one way and the other in the opposite direction. BiLSTMs effectively improve the quantity of data available to the network, providing the algorithm with more context.

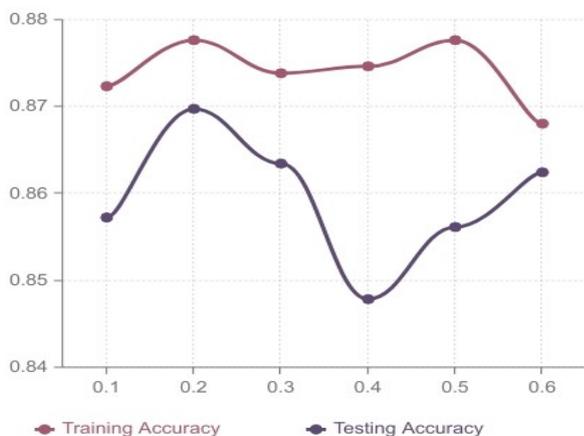
### 4 Results and Discussion

Dataset:

- Using the imdb reviews dataset from tfds(TensorFlow datasets) which contains 50,000.
- Convert data into a batch size of 64 for batch training.
- Perform text vectorization on the data to convert text data into vectors of numbers.

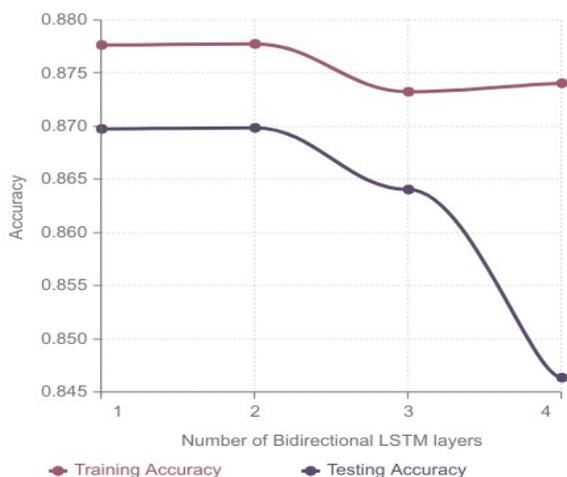
There are 8 layers in model architecture. The first layer is the text vectorization layer followed by an embedding layer. Then there are two pairs of Bidirectional and Dropout layers. Finally, there is a dense layers as the output layer.

**Effect of Dropout on Training and Testing Accuracy**



**Figure 4.1.** Effect of Dropout on Training and Testing Accuracy

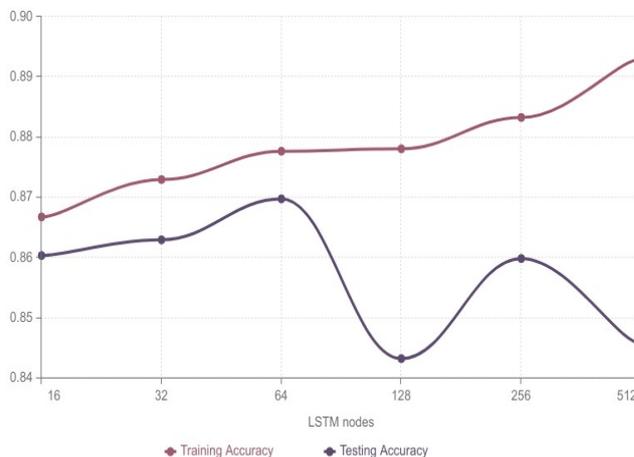
In Figure 4.1, the X-axis represents the dropout rate ranging from 0.1 to 0.6, and the Y-axis represents the accuracy. The accuracy becomes optimum at 0.2. Further training accuracy increases, but testing accuracy decreases due to over fitting. So 0.2 dropout rate is optimum.



**Figure 4.2.** Effect of Number of Bidirectional LSTM layers on Training and Testing Accuracy

In Figure 4.2, the X-axis represents the number of Bidirectional LSTM layers ranging from 1 to 4, and the Y-axis represents the accuracy. In this case the accuracy becomes optimum at 2 layers. On further increasing number of layers both training and testing accuracy decreases.

**Training and Testing Graph**



**Figure 4.3.** Effect of Number LSTM nodes on Training and Testing Accuracy

In Figure 4.3, the X-axis represents the number of LSTM nodes ranging from 16 to 512, and the Y-axis represents the accuracy. The accuracy becomes optimum at 64 nodes.

**Table 4.1.** Effect of Dataset on model Accuracy

Dataset	Training Accuracy	Testing Accuracy
IMDb Dataset	87.42%	86.56%
Rotten Tomatoes Dataset	82.53%	82.86%

As is evident from Table 4.1, it can be inferred that using the proposed model the training and testing accuracy on both the datasets are within the same range. Both the datasets are of different sizes, therefore the accuracy varies slightly. If two datasets of similar sizes are taken then one could get near the same accuracy.

**Table 4.2.** Effect of Text Vectorization on model Accuracy

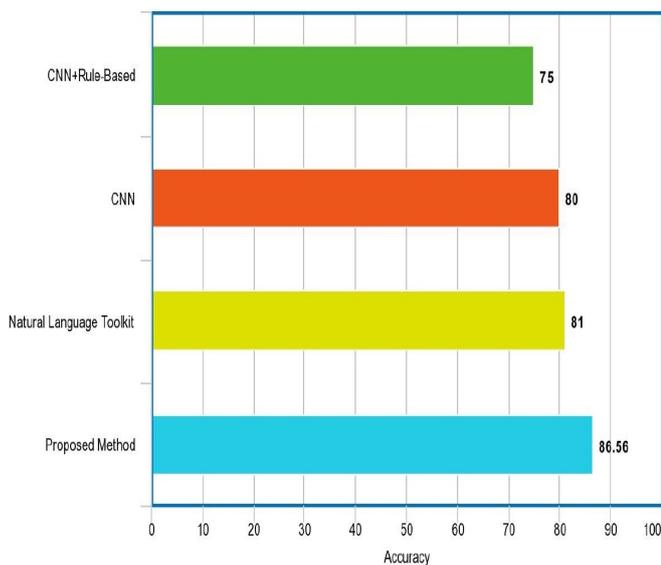
Text Vectorization	Training Accuracy	Testing Accuracy
tf-idf	87.42 %	86.56%
count Vectorizer	53.99 %	52.29 %
int	55 %	56.29%

As is evident from Table 4.2, tf-idf vectorizer gives the optimum training and testing accuracy.

**Table 4.3.** Effect of Activation Function on modal Accuracy

Activation Function	Training Accuracy	Testing Accuracy
Relu	87.42 %	86.56%
Sigmoid	87.42 %	85 %
Softmax	83.1 %	81.94%
Selu	88 %	86%
Elu	87.78 %	84.85%

The activation function specifies a neuron’s or node’s output in response to a single or multiple inputs. Selection of activation function is important for the performance of the model. As shown in Table 4.3, Softmax has slightly poor performance, otherwise all other activation function has similar performance, relu having the best training and testing accuracy.



**Figure 4.4.** Effects of different systems on Accuracy

Upon testing on multiple sets of hyperparameters ,it was found that 64 LSTM nodes,2 Bi-directional Layers and a

0.2 Dropout rate provides optimum accuracy. Tf-idf text vectorization and ReLU activation function provides best accuracy. Using these parameters the proposed model has the testing accuracy of 86.56% while existing systems using CNN provides 80% accuracy. Existing systems using Natural Language Toolkit achieves an accuracy of 81%.

## 5 Conclusion

The proposed binary classification model could rate comments as positive or negative using Bi-LSTM. For optimal accuracy 64 LSTM nodes,2 Bi-directional Layers and a 0.2 Dropout rate should be used. Tf-idf text vectorization along with ReLU activation function provides best results.Accuracy of the proposed system is better than of CNN and Natural Language Toolkit. The model was tested on two different datasets and achieved comparable results.

## References

- [1] S.M.Qaisar, “Sentiment Analysis of IMDb Movie Reviews Using Long Sort Term Memory”, 2020, *College of Engineering, Effat University, 21478, Jeddah, Saudi Arabia, IEEE*.
- [2] A.Sajeevan & Lakshmi.K.S, “An enhanced approach for movie review analysis using deep learning techniques”, *Proceedings of the Fourth International Conference on Communication and Electronics Systems (ICCES 2019)*, pp. 1788-1794.
- [3] Guixian Xu, Yueting Meng, Xiaoyu Qiu, Ziheng Yu & Xu Wu, “Sentiment Analysis of Comment Text Based on BiLSTM”, *IEEE Access*, vol. 7, Mar 31. 2019, pp. 51522-51532.
- [4] A.U.Rehman, Ahmad K.Malik, B.Raza & W.Ali, “A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis”, 2019, *Springer*.
- [5] G.P.S.Varma, A.Govardhan & I.Hemalatha, “Sentiment Analysis Tool using Machine Learning Algorithms”, *Elixir Comp. Sci. & Engg.* 58, *Elixir International Journal*, Jul. 2013, pp. 14791-14794.
- [6] M.S.Solanki, “Sentiment Analysis of Text using Rule Based and Natural language Toolkit”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *Blue Eyes Intelligence Engineering & Sciences Publication*, vol. 8, Oct. 2019, pp. 164-168.
- [7] A.Chakrabarti, “A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis” *Applied Computing and Informatics*, *Emerald Publishing Limited*, Feb. 2019, pp. 163-178.
- [8] P.Srivastava, *Essentials of Deep Learning: Introduction to Long Short Term Memory*, Dec 10, 2017. Accessed on: Feb 20, 2021. [Online]. Available: <https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/>
- [9] R.Wolff, "Sentiment Analysis & Machine Learning", April 20, 2020. Accessed on: Feb 24, 2021. [Online].

- Available: <https://monkeylearn.com/blog/sentiment-analysis-machine-learning/>
- [10] C.Olah, "Understanding LSTM Networks", Aug 27, 2015. Accessed on: Mar 15, 2021. [Online]. Available:<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [11] Shiao-li Green, "Sentiment Analysis — A how-to guide with movie reviews", Towards Data Science, Dec 20, 2019. Accessed on: Mar 20, 2021. [Online]. Available: <https://towardsdatascience.com/sentiment-analysis-a-how-to-guide-with-movie-reviews-9ae335e6bcb2>