PERFORMANCE OF MULTILAYER PERCEPTRON BASED SENTIMENT ANALYSIS ON HYPER PARAMETERS AND OPTIMIZERS

D P Gaikwad
Department of Computer Engineering
AISSM College of Engineering, Pune,

Abstract: **Sentiment analysis is a classification problem which is used to predict the polarity of words, paragraph or documents and then classify them into positive or negative sentiment. Sentiment analysis offers people a fast and effective way to measure the feelings towards product, their party and films etc. The primary issue in sentiment analysis techniques is the determination of the most appropriate classifier for a given classification problem. Several machine-learning techniques such as logistic regression, support vector machine and Naive Bayes have used to classify feelings of people. Now a day, deep learning is mostly used for sentiment analysis.**

In this paper, we propose the N-gram evaluation based Multilayer Perceptron as a sentiment analyser. The movie review dataset have been used for training and testing Multilayer Perceptron with different parameter. For a given dataset, the goal is to find the optimal parameters of Multilayer Perceptron to achieve maximum accuracy while minimizing computation time required for training and testing. In this experimental work, Multilayer Perceptron have trained and tested using different hyper parameters and different optimization functions. The experimental results show that the proposed N-gram evaluation based Multilayer Perceptron exhibits best training accuracy with minimum training time using Adam optimization function. It is also found that both RMSprop and Adam exhibits approximately same accuracy on test dataset.

***Index Terms* - N-gram, Machine Learning, Adam, Multilayer Perceptron, Deep learning, Dropout rate.**

I. INTRODUCTION

Sentiment analysis is the process of extracting emotions or opinions from a piece of text for a given topic [1]. Using sentiment analysis, we can understand the opinion, attitudes and emotions in the given text. It is also used to predict and analyze the hidden information present in the text. This hidden information is very useful to get insights of user's likes and dislikes. Sentiment analysis can also be applied to audio, images and videos [2]. Now a day, Internet has become the important part of every human life. Most of the people use online social networking sites, blogging sites, news articles and forums to express their opinions on certain things. These data sources can be used in stock markets, news articles or political debates in decision making. They also use these sites to know what other people's opinions are. Sentiment analysis technology is used to mine the knowledge from large volume of customer feedback, comments or reviews of any item, product or topic. It is very useful for new buyer to choose products or sellers in different sites. Sentiment analysis is also known as Opinion Mining, was first put forward by Pang Bo in 2002, which is the process of processing, analyzing, reasoning and inducing the subjective text with emotion[3]. In other words, Sentiment analysis is the task of detecting, extracting and classifying reviews, attitudes, and opinions concerning different topics. It is also called attitude analysis, appraisal extraction, opinion mining or review mining [4]. It plays an important role in Natural Language processing which is used to determine whether given text is positive, negative or neutral.

Document-level sentiment analysis is used to categorize an attitude text as conveying a positive or negative attitude from complete text. Sentence-level sentiment analysis is used to categorize sentiment conveyed in individual sentence. It will decide whether subjective sentence conveys positive or negative feelings. Aspect-level sentiment analysis is used to categorize the sentiment using feature of objects. The customer can give dissimilar opinions for dissimilar features of the same object. Aspect-level sentiment analysis uses sources of data such as the product reviews, news articles or political debates. Aspect-level sentiment analysis plays the essential role in the selection of business decision making by analyzing customers opinions [5]. Machine Learning Based approaches of sentiment analysis uses leaning algorithm which learns from past experience and helps in predictive analysis. Lexicon based approaches are based on sentiment vocabulary. In Hybrid Approaches, the collaboration of machine learning and lexicon is used to increase the performance. Concept Based Approaches is knowledge based. It analyzes the conceptual information associated to natural language opinion behind the multi-word expressions [6] [7] [8].

Sentiment analysis is a difficult task due to intricate and ambiguous nature of their expression. Therefore, automation of sentiment extraction from big or complex document is a big challenge. Public opinions are usually briefly written and jargon or idiom in nature with grammatical and typo mistakes. In such content, sentiment analysis or opinion mining is comparatively difficult than regular texts. Recently, machine leaning algorithms are mostly used to predict whether a document represents positive or negative opinion. There are two categories of machine learning one is supervised and other is unsupervised machine learning algorithms. Supervised algorithm uses a labelled dataset where each document of training set is labelled with appropriate sentiment class. Whereas, unsupervised learning include unlabeled dataset where text is not labelled with appropriate sentiments.

In this paper, we present the extension work of our previous paper which is used to for reviewing films. This paper presents the role of text pre-processing in sentiment analysis with appropriate feature selection. It also presents performance of the proposed multilayer perceptron model with different hyper parameters and optimization functions. The remainder of this paper is organized as follows. In Section 2, several related works are introduced. In Section 3, the proposed sentiment analysis model is described. In Section 4, the experimental results and analysis are presented. In Section 5, the whole paper is concluded.

II. RELATED WORK

Now a day, a lot of work has been done in the field of sentiment analysis on twitter, face book comments and film reviews. Several methods have adapted to automatic prediction of expression, sentiments of word or a document. The movie reviews have considered as a dataset for training different models. In this section, some contributions have been presented by number of researches.

Xian Fan et.al [9] has proposed sentiment lexicon based on word vectors. They have compared the effect of different sentiment lexicon on the performance of test sentiment analysis using classical naive Bayes. The results show that the lexicon built by word2vec for sentiment analysis gets the higher precision and recall rate. Neethu M S and Rajasree R [10] have proposed different algorithms such as SVM, Naive Bayes and some ensemble classifiers for sentiment analysis. The accuracy of all these classifiers has almost similar accuracy. The accuracy of SVM and Ensemble classifier has similar recall and accuracy. The precision value of Naive Bayes has better compared to the SVM ensemble classifiers. They have obtained maximum accuracy of 90% using SVM and Ensemble classifiers. Go and L. Huang [11] have implemented Naive Bayes, MaxEnt and Support Vector machine for sentiment analysis for twitter data. They have used feature space consisted of unigrams, bigrams and POS. They concluded that unigram, bigram and POS are more effective as features. They observed that SVM outperformed other models. Turney et al [12] have used bag-of-words method for sentiment analysis. In this method, the relationships between words were not considered. A document is represented as just a collection of words. They have used aggregation function to unite sentiments of every word to determine the sentiment for the whole document. Siyuan Chen and Chao Peng [13] have proposed a deep neural network model combining convolutional neural network and regional long short-term memory for sentiment analysis. The experimental results show that the proposed model performs better than SVM, attention based LSTM and multi-attention based CNN. The proposed model also take less time for training the model.

Vishal A. Kharde and S.S. Sonawane [14] have provided a survey on the performances of Naive Bayes, Max Entropy and Support Vector Machine used for opinion mining. The experimental results show that SVM and Naive Bayes have the highest accuracy. They concluded that lexicon-based methods are very effective in some cases. They have suggested that bigram model provides better sentiment as compared to other models. Kai Yang et. Al. [15] has constructed the domain sentiment dictionary using external textual data. They have proposed a hybrid model combining SVM and Gradient Boosting Decision Tree GBDT together. The accuracy of GBDT is higher than that of SVM by 3-4%. They also observed that SVM performed well on simple structured sentences and strong opinion tendency. SVM has poor performance for those complicated sentences. Gradient Boosting Decision Tree performs well for long sentences with many sentiment words.

Yuling Chen and Zhi Zhang [16] have combined Convolutional Neural Networks (CNNs) with SVM for text sentiment analysis. In this proposal they have used CNN as an automatic feature learner and SVM as a classifier. The proposed combined method improved the accuracy of text sentiment classification effectively compared with traditional CNN. Amolik and M.Venkatesan [17] have implemented two machine learning classifiers. Both classifiers performed well and also provide higher accuracy. The results show that 75 % accuracy got form SVM and 65% accuracy form Naïve Bayesian classifier. By increasing the training data, the accuracy of classification can be increased. Nagamma P and Pruthvi H.R [18] have proposed clustering algorithm which resulted in good accuracy. They concluded that clustering with a classification model reduces the data used for prediction. It could be seen predominantly if the dataset used is large.

.

III. METHODOLOGY AND PROPOSED SYSTEM

In this study, the Internet movie database movie review dataset is used which contains movie reviews posted by people on the IMDb website for sentiment analysis. It contains positive or negative corresponding labels which indicate whether the reviewer liked the movie or not. Initially, dataset is transformed to a format that can understand by model. The training and testing datasets are converted into numerical vectors. Firstly, texts in dataset are divided into words or smaller sub-texts to determine the set of unique tokens present in dataset. This process is called tokenization to create “Vocabulary” of dataset. Then these token are converted in N-gram vectors to represent each token.

In n-gram vectorization, text is represented as a collection of unique n-grams (group n adjacent tokens). In this work, word unigrams + bigrams tokenising have used to provide good accuracy with less compute time. Then indexes are assigned to the unigrams and bigrams. These indexed n-grams are converted into to numerical vectors which can process by machine learning models. The f_classif function is used to select 20,000 important features or token from tens of thousands of tokens.

Machine learning Models are broadly classified into two categories: sequence model and other n-gram model. Sequence model use word ordering information and n-gram model just see text as “bags” (sets) of words. The convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variations are examples of sequence models.

The support vector machine, logistic regression , gradient boosted trees, simple multilayer perceptron, fully-connected neural networks are example of n-gram machine learning models. Multiplayer perceptron is very simple to define and understand. This model also provides good accuracy and takes relatively little time to train.
In this paper, the N-gram based MLP machine learning model has been used for sentiment analysis. Figure 1 summarises the process of proposed sentiment analysis system.

Our proposed system consists of mainly following six steps.

1. Calculate the number of samples/number of words per sample ratio.
2. Tokenize the text as n-grams.
3. Split the samples into word n-grams.
4. Convert the n-grams into vectors.
5. Score the importance of the vectors and then select the top 20K using the scores.
6. Build an MLP model.
7. Measure the model performance with different hyper parameter values to find the best model configuration for the dataset.

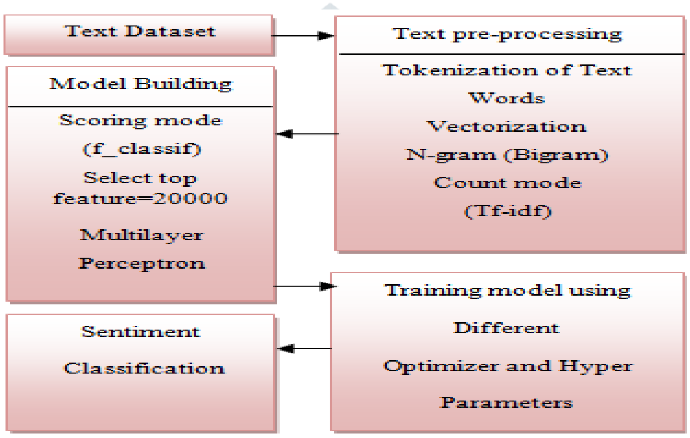


Figure.1. Sentiment Analysis Process

.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the results of several experiments to assess the performance of the classifier have reported. All experiments have performed on Lenovo laptop with configurations: Processor: Intel(R) Core(TM) i5 4300U @1.9GHz, RAM: 4.0GB, 64-bit operating system, x64-based processor. The model has trained using different hyper parameters such as number **of layers in the model, Learning rate, dropout rate and number of epochs. In first phase, we** have studied the influence of different hyper parameters on performance of sentiment analysis for long texts. The performance of the sentiment analysis model is measured based on the accuracy on the dataset labelled as positive and negative. Several experiments have conducted using training dataset of 25000 samples and testing dataset of 25000 samples. The different hyper parameters have been used for experimentation purpose. Figure 2 show the structure of proposed sentiment analysis model.

```
Layer (type)                 Output Shape              Param #
=================================================================
dropout (Dropout)            (None, 20000)             0
_________________________________________________________________
dense (Dense)                (None, 256)               5120256
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 257
=================================================================
Total params: 5,120,513
Trainable params: 5,120,513
Non-trainable params: 0
_________________________________________________________________
```

Figure.2. Structure of the proposed Sentiment Analysis model.

Finally, the model has trained using all optimized parameters. Table 1 compares the performance of N-gram evaluation based multilayer perceptron model in term of accuracy and time required to train model.

Table 1. Accuracy and time required for training model.

Sr. No.	Common Parameters	Training Accuracy In %	Accuracy on test dataset	Approx Time require to train model in Min.
1	Lr=0.0001, Epochs=200, Batch size=256, Layers=2 Units=256, Dropout rate=0.4	95.44	90.63%	23.33
2	Lr=0.0001, Epochs=200, Batch size=512, Layer=2, Units=256, Dropout rate=0.4	95.58	90.67%	27.53
3	Lr=0.0001, Epochs=100, Batch size=512, Layer=2, Units=2, Dropout rate=0.4	95.87	90.68%	30.0

Following figures show the accuracy and model loss on training and test dataset. Figure 3, figure 5 and figure 7 show the accuracies of different model with different hyper parameters. Figure 4, figure 6 and figure 8 show the loss on training and test dataset.

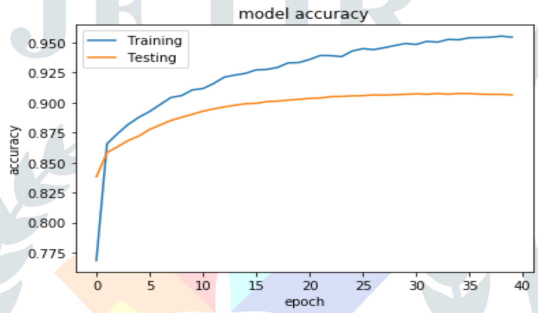


Figure 3. Accuracy of model 1 on training and testing dataset.

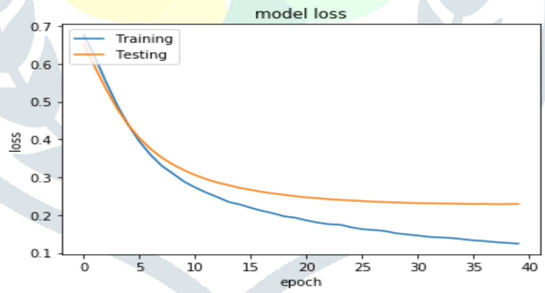
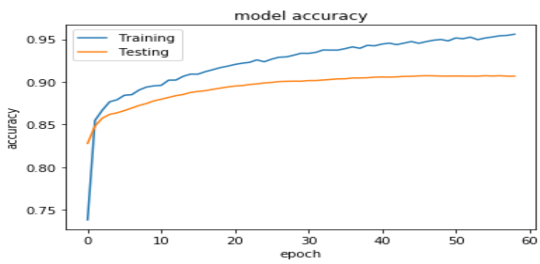


Figure 4. Model loss of model 1 on training and testing dataset.



I.
Figure 5. Accuracy of model 2 on training and testing dataset.

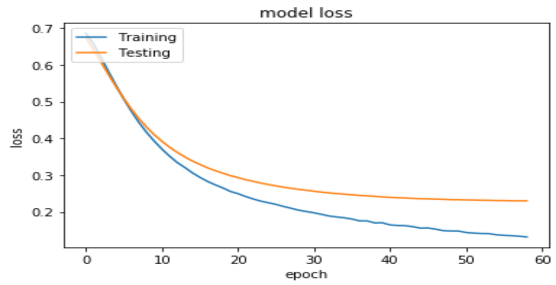


Figure 6. Model loss of model 2 on training and testing dataset.

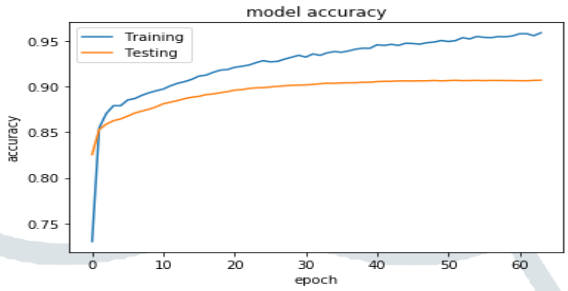


Figure 7. Accuracy of model 3 on training and testing dataset.

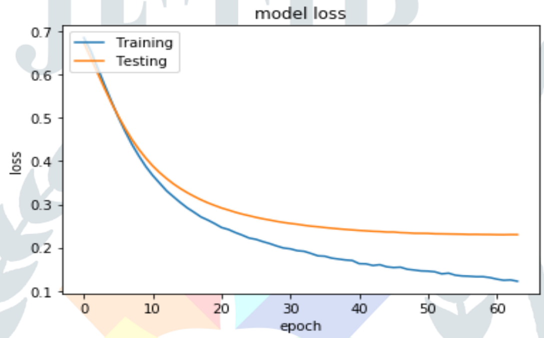


Figure 8. Model loss of model 3 on training and testing dataset.

In second phase, the proposed model has been trained using different optimization functions to study the influence of them on performance of sentiment analysis model. Optimization function that decides how the network weights will be updated based on the output of the loss function. The proposed model has trained using following optimization functions.

SGD: Stochastic Gradient Descent optimizer support for momentum, learning rate decay and Nesterov momentum.

RMSprop: This optimizer is usually a good choice for recurrent neural networks.

Adagrad: It is an algorithm for gradient-based optimization that greatly improved the robustness of SGD and used it for training large-scale neural networks.

Adadelta: Adadelta is a more robust extension of Adagrad that adapts learning rates based on a moving window of gradient updates, instead of accumulating all past gradients. It continues learning even when many updates have been done. In Adadelta optimizer, you don't have to set an initial learning rate.

Adam: Adaptive Moment Estimation is another method that computes adaptive learning rates for each parameter which keeps an exponentially decaying average of past gradients.

Adamax: It is a variant of Adam based on the infinity norm.

Nadam: Much like Adam and it is essentially RMSprop with momentum; Nadam is Adam RMSprop with Nesterov momentum.

All experiments have conducted using training dataset of 25000 samples and testing dataset of 25000 samples. Table 2 and Table 3 compares the performance of N-gram evaluation based multilayer perceptron model in term of accuracy and training using different optimization functions. Figure 9 to figure 16 shows the accuracy of models on training and test dataset with different optimization functions.

Table 2. Accuracy of models on optimizers.

Sr. No.	Hyper Parameters	Optimizer	Accuracy on Training dataset	Accuracy on Test dataset

1	Lr=0.0001, Epochs=100, Batch size=512, Layer=2, Units=256 Dropout rate=0.4	SGD	50.18	50.18
2		RMSprop	94.62	90.83
3		Adagrad	87.79	86.09
4		Adadelta	51.37	52.06
5		Adam	96.00	90.62
6		Adamax	92.52	90.37
7		Nadam	95.86	90.63

Table 3. Training time of model on optimizers.

Sr. No.	Hyper Parameters	Optimizer	Approx. time to train model in min.
1	Lr=0.0001, Epochs=100, Batch size=512, Layer=2, Units=256 Dropout rate=0.4	SGD	40 min.
2		RMSprop	30 min.
3		Adagrad	43min
4		Adadelta	50min.
5		Adam	30 min.
6		Adamax	47min.
7		Nadam	32min.

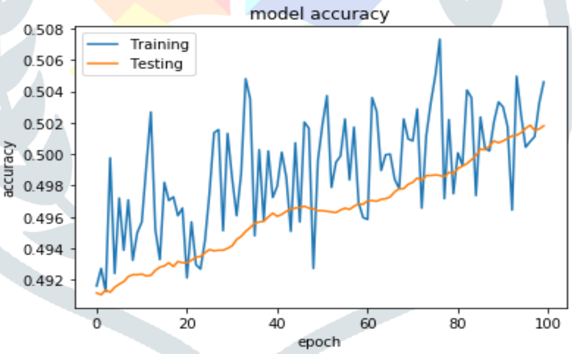


Figure 9. Accuracy on training and test dataset using SGD.

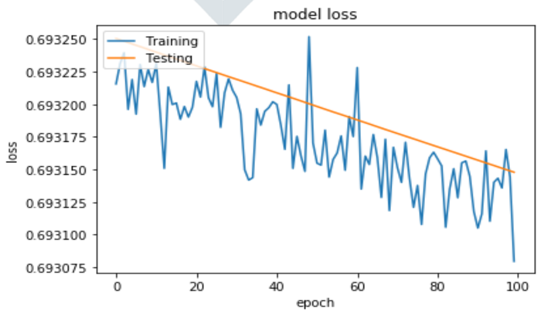


Figure 10. Loss on training and test dataset using SGD.

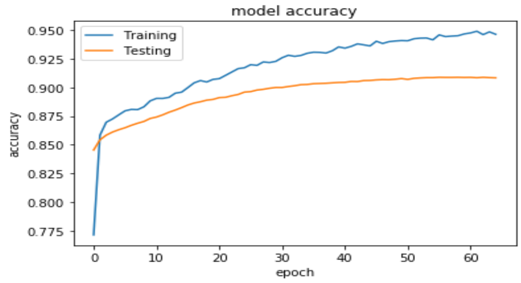


Figure 11. Accuracy on training and test dataset using RMSprop.

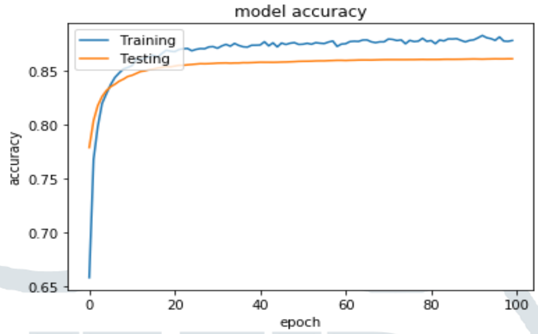


Figure 12. Accuracy on training and test dataset using Adagrad.

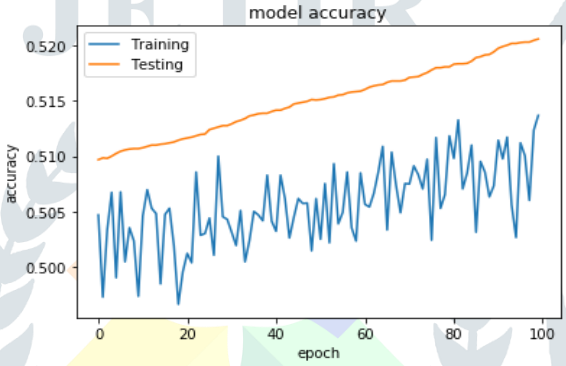


Figure 13. Accuracy on training and test dataset using Adadelta.

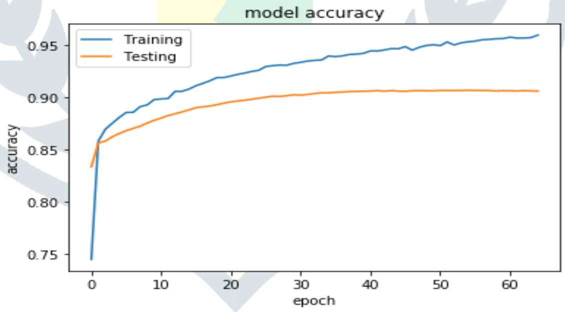


Figure 14. Accuracy on training and test dataset using Adam.

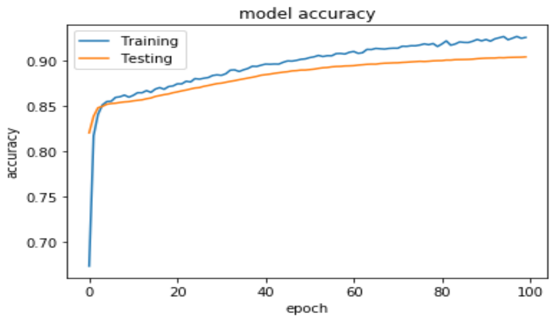


Figure 15. Accuracy on training and test dataset using Adamax.

Figure 16. Accuracy on training and test dataset using Nadam.

V. CONCLUSIONS AND FUTURE WORK

In this paper, the importance of sentiment analysis system has studied in detail. The sentiment analysis system has implemented using Multilayer perceptron. The pre-processing of long text have described in detail. For training the model, the IMDb Database has been used. It contains movie reviews posted by people on the IMDb website for sentiment analysis. The proposed N-gram evaluation based machine learning model has trained using different hyper parameters and optimization functions. In this experimental study, an attempt has been made to tune different hyper parameters of N-gram based Multilayer perceptron to provide maximum accuracy with minimum training time. The models have trained using training dataset of 25000 samples and tested on 25000 samples. The models have trained on all optimized hyper parameters. The experimental results have studied and display in this paper. The experimental results show that the proposed N-gram evaluation based Multilayer Perceptron exhibits best training accuracy with minimum training time using Adam optimization function. It is also found that both RMSprop and Adam exhibits approximately same accuracy on test dataset. In future, sentiment analysis system will be implemented using sequential learning model to improve the correlation ship of sentences in the document which can help to give opinion on very long text dataset.

REFERENCES

[1] Pang B, Lee L., "Opinion mining and sentiment analysis" Found Trends Inform Retriev:1–135, 2008.

[2] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey" Ain Shams Engineering Journal 5.4 :1093-1113, 2014

[3] Pang, Bo, L. Lee, and S. Vaithyanathan, "Sentiment classification using machine learning techniques," Computer Science, pp.79—86, 2002.

[4] Ravi, K., & Ravi, V., "A survey on opinion mining and sentiment analysis: Tasks, approaches and application," Knowledge-Based Systems, 89, pp. 14-46, 2015.

[5] Penubaka Balaji, D.Haritha and O.Nagaraju, "An Overview on Opinion Mining Techniques and Sentiment Analysis," International Journal of Pure and Applied Mathematics, Applied Mathematics Vol. 118, No. 19, ISSN: 1311-8080, pp. 61-69, 2018.

[6] Akashi Kumar, "Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence," Webology, Volume 14, Number 2, December, 2017.

[7] Medhat W., Hassan, A. and Korashy, H., "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, Vol. 5, No. 4, pp. 1093-1113, 2014.

[8] Cambria, E., Schuller, B., Xia, Y. and Havasi, C., "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, Vol. 28, No. 2, 15-21, 2013.

[9] Xian Fan et.al., "Apply Word Vectors for Sentiment Analysis of APP Reviews," In The 2016 3rd International Conference on Systems and Informatics (ICSAI), pp. 1062-1066, 2016.

[10] Neethu, M. S., & Rajasree, R., "Sentiment analysis in twitter using machine learning techniques," In Computing, Communications and Networking Technologies (ICCCNT), Fourth International Conference on (pp. 1-5). IEEE, July 2013.

[11] Go, R. Bhayani, L.Huang, "Twitter Sentiment Classification Using Distant Supervision," Stanford University, Technical Paper, 2009.

[12] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417–424, Association for Computational Linguistics, 2002.

[13] Siyuan Chen and Chao Peng, "A Deep Neural Network Model for Target-based Sentiment Analysis," In 2018 International Joint Conference on Neural Networks (IJCNN), 2018.

[14] Vishal A. Kharde and S.S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016.

[15] Kai Yang et. al., "An Effective Hybrid Model for Opinion Mining and Sentiment Analysis," BigComp 2017, 978-1-5090-3015-6/17/$31.00 ©2017 IEEE.

[16] Yuling Chen and Zhi Zhang, "Research on text sentiment analysis based on CNNs and SVM," Wuhan University of Science and Technology, 978-1-5386-3758-6/18/$31.00_c IEEE, 2018.
[17] Amolik, A., Jivane, N., Bhandari, M., and Venkatesan, M., "Twitter sentiment analysis of movie reviews using machine learning techniques," International Journal of Engineering and Technology, Vol. 7, No.6, 2038-2044, 2015.
[18] Nagamma, P., Pruthvi, H. R., Nisha, K. K., and Shwetha, N.H, "An improved sentiment analysis of online movie reviews based on clustering for box-office prediction," In Computing, Communication & Automation (ICCCA), International Conference on (pp. 933- 937).IEEE, May 2015.