A COMPARATIVE STUDY ON SUPERVISED LEARNING METHODS

Mr. Sunil M E[1], Dr. Vinay S[2], Mr. Ramakrishna Vaidya[3]
[1]Computer Science & Engineering*(VTU)*, PESITM, Shivamogga, India
[2]Information Science & Engineering (VTU), PESCE, Mandya,India
[3]Computer Science & Engineering*(VTU)*, PESITM, Shivamogga, India

Abstract : Web has become the part of human daily life and continuously generates the sentiment in the form of feedbacks, reviews. The analyzing and classifying these sentiments is a tedious task. In this paper we are focusing on measuring the classification accuracy of supervised learning algorithms like RNN, SVM, NB and CNN on IMDB movie reviews.

IndexTerms **- Sentiment Classification, Movie Review, Supervised leaning, Recall, Precision, IMDB.**

I. INTRODUCTION

The era of Internet produces huge amount of information created and consumed by so many users daily. Information has been growing rapidly leading to new arenas in which user can express their opinions through blogs, websites, social media like Facebook, Twitter etc. People are open to express their opinion about particular topic without hesitating and influenced by others. This will help others to arrive into a decision and also helps the service provider to understand the user demands, likes and dislikes and also improve the services by taking appropriate actions. Sentimental analysis is a process of expressing feeling, attitude, judgment or thought [1]. It is also called as opinion mining. There are different parts in sentimental analysis: document view, sentence view and aspect view [2]. One such domain of review is Movie reviews in which viewers are free to express their feedback towards a particular movie. This affects everyone including audience, production companies and other stakeholders. The sentiments expressed in movie reviews are informal and free from grammar but they give true reflection about the emotion that is being conveyed. Classification of review sentiment requires a method to extract the meaning full information from feedbacks and it requires a critical analysis of all the algorithms and process required to decide the sentiment. In this paper we are trying to evaluate the accuracy of few machine learning algorithms such as Naive Bayes, recurrent neural network, support vector machines, and convolution neural network on IMDB movie reviews.

The rest of this paper is organized into six main parts. Section 2 presents the basic process of sentimental analysis and sentiment related literature in field of machine learning on IMDB reviews. Section 3 explains the details of IMDB review sentiment dataset. Section 4 gives the brief over view of algorithms used in this study. Section 5 describes the measuring parameters and Section 6 discusses the comparison and results of the experimental study.

II. LITERATURE REVIEW

A sentimental analysis is a computational study of people's emotion toward an entity. Entity can be individuals, topics or events. Opinion mining is synonyms of sentimental analysis. Sentimental analysis is the process of finding opinion, identify the sentiment and classify the polarity [3]. Bing Liu [4] defines as "opinion is a quintuple <oi, fij, soijkl, hi, tl>, where oi is the target object, fij is the feature of the target object oi, hi is the opinion holder, tl is the time when the opinion is expressed and soijkl is the sentiment value of the opinion expressed by the opinion holder hi about the object oi at time tl". Current research in sentiment analysis focuses on classifying sentiment according to their polarity as positive, negative or neutral. One Standard technique was developed by [5] which analyzed the movie reviews with different machine learning algorithms. This technique was the base for many researchers to work on text classification. In addition a comparative study on different classification approaches has been performed to understand which classifier suits for study [2]. The process involves sentiment identification, feature selection, sentiment classification and sentimental polarity. Beyza Çizmeci and Şule Gündüz Ögüdücü have proposed Factorization Machines approach helps to predict movie success of newly released movies by analyzing IMDb ratings in social media data and compare it to current studies with the promising results [6]. Kamil Topal and Gultekin Ozsoyoglu produced the emotion map for a movie on basis of reviewers score and their emotion content aggregated and projected on a movie [7].

III. IMDB DATASET

In this study we have used publically available IMDB Movie Reviews dataset [8]. It contains 50000 highly polar movie reviews for training the model. In which 25000 of the reviews are used for training and 25000 for testing as shown in Table 1.

Table 1 Split of dataset

	Positive	Negative
Training	12500	12500
Testing	12500	12500

The following fig.1. shows example dataset used for experiments.

	type	review	label	file
0	test	Once again Mr. Costner has dragged out a	neg	0_2.txt
1	test	This is an example of why the majority of	neg	10000_4.txt
2	test	First of all I hate those moronic rappers, v	neg	10001_1.txt
3	test	Not even the Beatles could write songs e	neg	10002_3.txt
4	test	Brass pictures (movies is not a fitting wor	neg	10003_3.txt
5	test	A funny thing happened to me while wat	neg	10004_2.txt
6	test	This German horror film has to be one of	neg	10005_2.txt
7	test	Being a long-time fan of Japanese film, I	neg	10006_2.txt

Figure 1-Example dataset

Tables II and III shows two reviews from the IMDB dataset comprising a positive and a negative review.

TABLE 2. NEGETIVE REVIEW

"A friend gave me this movie because she liked it. I decided I would finally watch it. It was soooooooo long. I kept waiting for the suspense to happen but it never did. I kept waiting for something to happen after the opening scenes, and it never did. I stopped the movie and came back later. I actually forced myself to watch the rest of it hoping it would get better. It got worse. I kept asking myself, who are these people? Do they have feelings? are they just robots? I'm glad I didn't pay to see it or pay to rent it. The end would have been better if Dutch died from the gunshot wound. At least we would have gotten some emotion from the audience. Or maybe not."

TABLE 3. POSITIVE REVIEW

"I first saw this in the 70s on syndicated TV and admired its production values, which were high tech for the time. The remastered video is rich and colorful, far more intense then the pale 35 mm TV prints. This movie deserves more attention: it paved the way for UFO, Space: 1999 and even Star Wars with its detailed miniatures and cleverly conceived gadgets. Sure, the story of an alternative anti-matter planet Earth has been recycled a hundred times since Star Trek, but the beauty of this film is its self-conscious European flair for design: from the Rolls Royce space engines to the "Euro Sec" letterhead business paper, JFSS or Dopplegangers as it was called in Europe is enjoyable for the imaginary vision of Europe in space in the shadow of the Superpowers. Gerry and Sylvia Anderson's ambitious epic gets a little tedious when the American astronaut finally realizes that he is on the doppleganger Earth, and everything is literally downhill after the poetically graceful shuttle boarding sequence. A mediocre story is helped along by a grand and lyrical classical score by the late great Barry Gray, the John Williams of Britain."

IV. EXPERIMENT AND EVALUATION

In our experiment, we have used the four supervised classification algorithms namely RNN- LSTM, Support vector machine, Naïve Bayes and Convolution neural network. We have trained these classifiers using IMDb movie reviews to observe the classification accuracy of classifiers. We also calculated positive, negative, micro average, Macro average and weighted average for Recall, Precision, F1-score and support of each classifiers.

1.1 RNN with LSTM

Recurrent Neural Network is a power full neural network with long short term memory. RNN are able to remember the important information about the input they have received. It enables the RNN to predict what is coming next. RNN makes a decision on current input based on what it has learned from the previous input. LSTM are extension of RNN with extended memory, so LSTM are used as building units for the layer of RNN [9].

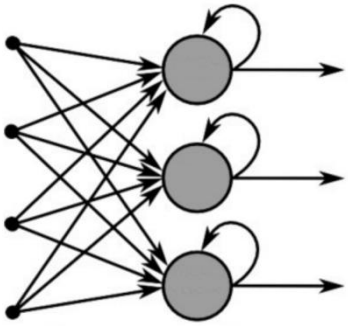

Figure 2 RNN

LSTM has input, output and Forget gate. Input gate helps to determine whether input is received or not, forget gate is used to delete the information which is not important and output gate impact the output of the current time step. The RNN with three gates has shown in fig.3 [10].

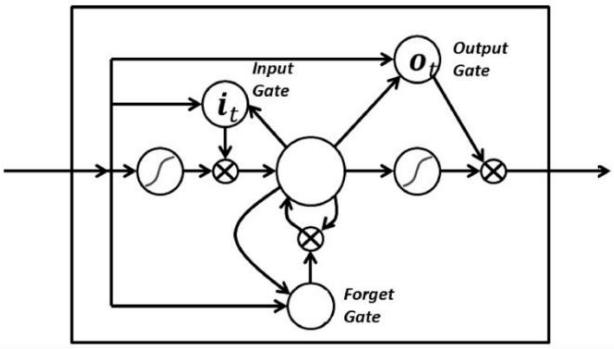

Figure 3 RNN with 3 gates

4.2 SVM with linear Kernal

In 1995 Vapnik proposed Support vector machine as a statistical classification method. Given a training examples, Possibly SVM generates separate hyper surface that has maximum generalization ability. SVM kernels are set of mathematical functions like linear, Sigmoid, Polynomial etc. In this paper we considered the linear kernel it can be used as normal dot product of any two given observations. The product between two vectors is the sum of the multiplication of each pair of input values. Fig .4 shows the SVM [12].

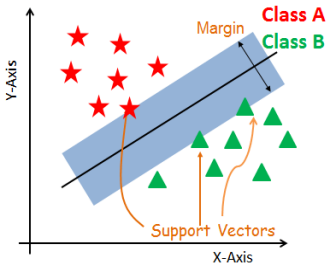

Figure 4-SVM

4.3 Naïve Bayes

Naïve Bayes is a probabilistic classifier based on Bayes theorem P (A|B) = (P (A) P (B|A))/P (B) where P (A) is the prior density and P(B|A) is the sample likelihood and P(B). Bayesian classifiers find the most likely class to a given examples feature vector. Assuming that features are independent from given class significantly simplified by learning this type of classifiers [13]. Naive Bayes is considered as an effective method in many practical applications, including medical diagnosis, text classification, and systems performance management [14].

4.4 Convolution Neural Network

Convolution Neural Network is a specific type of artificial neural network that can take input as any spatial data. It is also a part of deep learning which take the image as input. It assigns learnable weights and biases to various aspects and objects. Convolution operation is applied to input and passing the result to the next layer by Convolution layer [3]. Each Convolution neuron process the data only with respect its field.

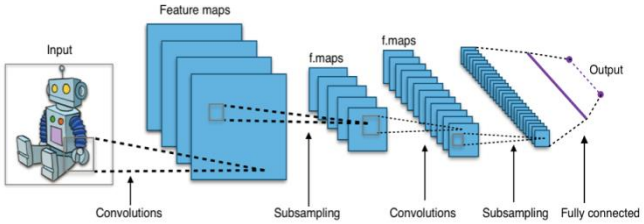

Figure 5-CNN

V. MEASURING PARAMETERS

In this paper we have measured the precision, recall, F1-Score and support of each algorithms and also compared the accuracy of SVM, NB, CNN, RNN for IMDb data.

The main idea is to determine the model that minimizes the overall cost, or maximize the overall gain of evaluated model for a given application [14]. Here we have to consider the confusion matrix with four outcomes: True Positive (TP) type I elements correctly classified, True Negative (TN) type II elements correctly classified, False Positive (FP) type II elements wrongly classified and False Negative (FN) type I elements wrongly classified [15].

5.1 Precision

Precision is the positive predictive value in the field of information retrieval. It is the fraction of relevant instances among the retrieved instances. It also states that the degree in which identified as type I elements belongs to type the type I class

$$\text{Precision} = TP / (TP+FP)$$

5.2 Recall

It is also known as sensitivity is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. In other words percentage of type I elements that the classifier manages to classify correctly. It can be interpreted as classifier's effectiveness.

$$\text{Recall} = TP / (TP+FN)$$

5.3 F-Score

F-Score is also called F1-score or F-measure is a measure of test the accuracy. It considers recall and precision to calculate the score. It also stated as a harmonic mean between the precision and recall.

$$\text{F-score} = (2*\text{Precision}* \text{Recall}) / (\text{Precision}+\text{Recall})$$

5.4 Accuracy

Over all percentage of correctly classified elements

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$$

5.5 Evaluation and Results

RNN-LSTM has been performed using the tensor flow with two Epochs and 25000 positive and negative reviews which produced the 86.81% accuracy. SVM with linear kernel and NB also performed on IMDb data in tensor flow give an 86.5% accuracy and 86.69% accuracy respectively. CNN has been performed using the tensor flow with two Epochs with 87.85% accuracy.

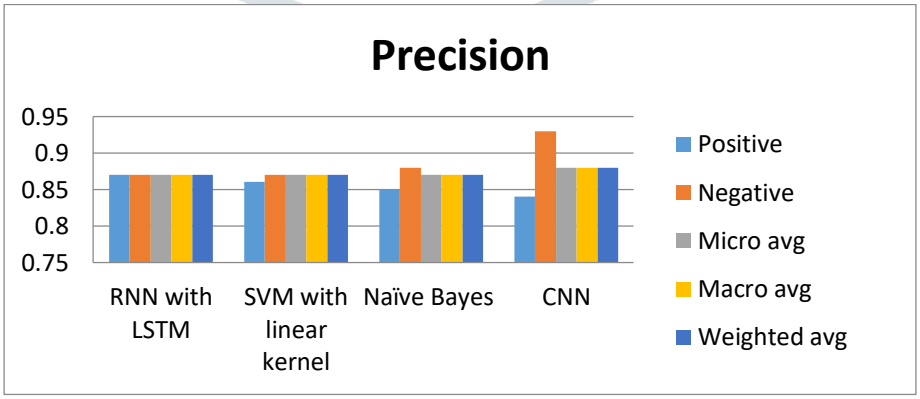

Figure 6- Precision of four algorithms

Fig.6. represents the precision of all algorithms versus positive, negative, micro average, macro average and weighted average.

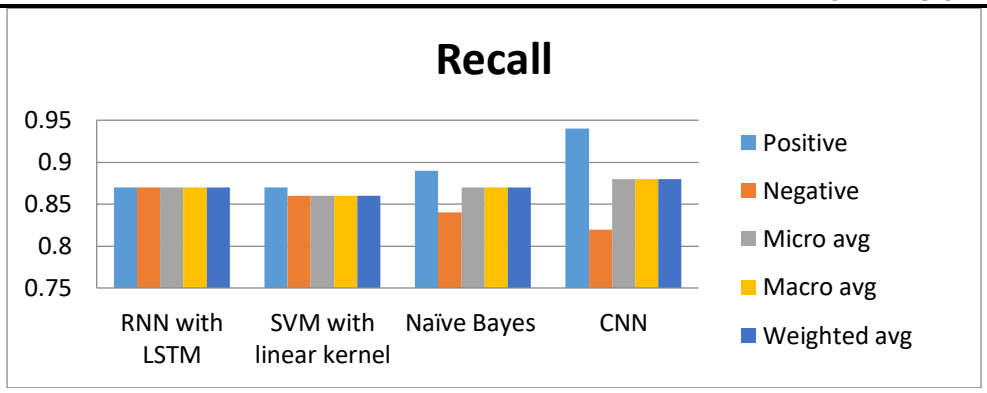

Figure 7- Recall of four algorithms

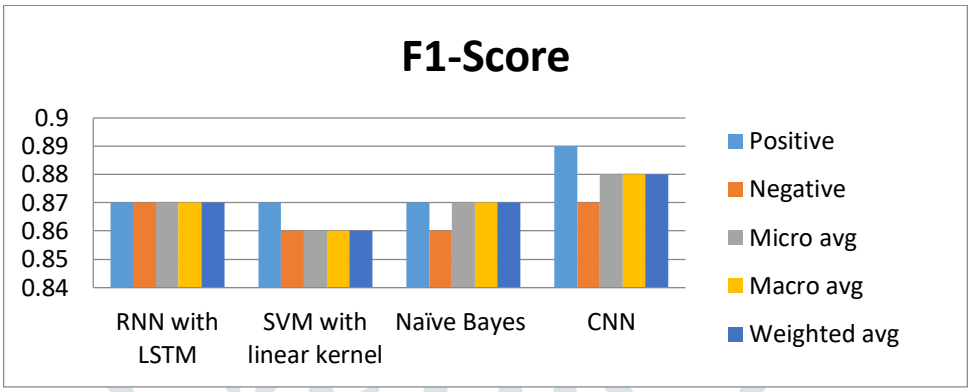

Figure 8- F-score of four algorithms

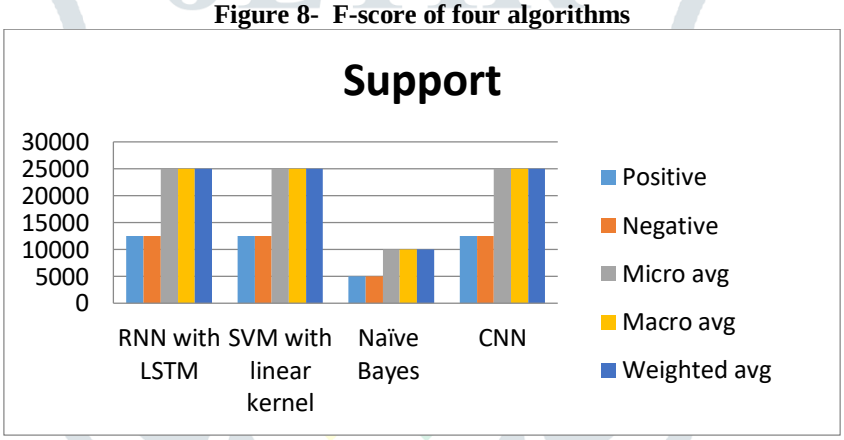

Figure 9- Support of RNN,SVM,NB and CNN

TABLE 4 COMPARISON BETWEEN DIFFERENT ALGORITHMS FOR IMDB POSITIVE

Methods	Accuracy	Precision	Recall	F-score
RNN with LSTM	86.81	87	87	87
SVM with linear kernel	86.5	87	86	86
Naïve Bayes	86.69	88	84	86
CNN	87.85%	84	94	89

TABLE 5 COMPARISON BETWEEN DIFFERENT ALGORITHMS FOR IMDB NEGATIVE

Methods	Accuracy	Precision	Recall	F-score
RNN with LSTM	86.81	87	87	87
SVM with linear kernel	86.5	86	87	87
Naïve Bayes	86.69	85	89	87
CNN	87.85%	93	82	87

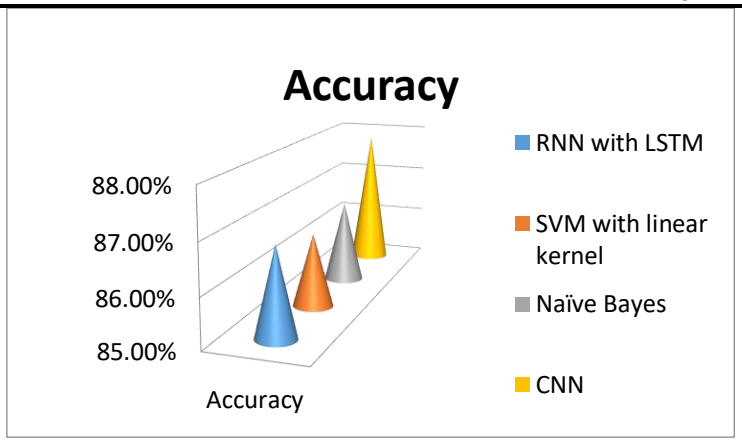

Figure 9-Accuracy

Fig.9 represents the rate of accuracy recorded by five Algorithms which are Recurrent Neural Network using LSTM, Support vector machine with linear kernel, Naïve Bayes and CNN.

VI. CONCLUSION

Our methodology suggests a way to analyze the IMBb movie reviews and comparing the accuracy of various supervised classification algorithms. Our results showed that all four algorithms considered for study gives a near accuracy. We can conclude that all algorithm gives similar accuracy. In future we plan to work on introducing variations to existing methods at different stages of sentimental analysis process to enhance the accuracy.

REFERENCES

[1] T. P. Sahu and S. Ahuja, "Sentiment analysis of movie reviews: A study on feature selection & classification algorithms," *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, Durgapur, 2016, pp. 1-6

[2] A. Yenter and A. Verma, "Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis," *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, New York City, NY, 2017, pp. 540-546.

[3] Walaa Medhat, Ahmed Hassan, Hoda Korashy, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*,Volume 5, Issue 4,2014

[4] Bing Liu, "Sentiment analysis and opinion mining", *Morgan and Claypool* , 2012.

[5] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?:sentiment classification using machine learning techniques." *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*-Volume 10. Association for Computational Linguistics, 2002.

[6] B. Çizmeci and Ş. G. Ögüdücü, "Predicting IMDb Ratings of Pre-release Movies with Factorization Machines Using Social Media," *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, Sarajevo, 2018, pp. 173-178.

[7] K. Topal and G. Ozsoyoglu, "Movie review analysis: Emotion analysis of IMDb movie reviews," *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, 2016, pp. 1170-1176.

[8] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts, "Learning Word Vectors for Sentiment Analysis". *The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).*

[9] *Alex Sherstinsky*."Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network". *CoRR abs/1808.03314 (2018)*

[10] *https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5*

[11] V. Vapnik, "The Nature of Statistical Learning Theory", *Springer, New York, 1995.*

[12] https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python

[13] I Rish "An empirical study of the naive Bayes classifier", *T.J. Watson Research Center*, IJCAI 2001

[14] Tom M. Mitchell. "Machine Learning". *McGraw-Hill*, 1997

[15] Tom M. Mitchell. "Machine Learning Juan D. Velásquez, Vasile Palade, Lakhmi C. Jain, "Advanced Techniques in Web Intelligence-2", *Studies in Computational Intelligence book series* (SCI, volume 452)