

# Effective Deep Learning Model using Twitter Records

Akshay Kumar Yadav

*M.Tech in Computer Science & Engineering in Big  
Data Analytics, School of Computer  
Science and Engineering (SCOPE), VIT Chennai,  
Chennai, Tamil Nadu, India.  
Email: akshaykumar.yadav2020@vitstudent.ac.in*

Dr. V Vani

*Assistant Professor (Sr.)  
School of Computer Science and Engineering  
(SCOPE) VIT Chennai,  
Chennai, Tamil Nadu, India.  
Email: Vani.v@vit.ac.in*

**Abstract** - Now a day's sentimental analysis is widely used to understand and interpret the opinion, speculations and reaction of the population. Sentimental analysis helps in identifying the reaction of people towards a topic, person or brand which helps in improving brand image, updating strategies, and knowing the customers' demands and needs. The objective of this survey paper is to perform sentimental analysis using social media platform twitter.

**Keyword** – Sentimental Analysis, Twitter, Opinion, Speculation

## I. INTRODUCTION

In today's dynamic era with the increased use of internet information is easily accessible from anywhere and anytime. We can use the feedbacks and reaction to understand the customer needs and their attitude towards products, people and different services. Sentimental analysis can be used at many levels, topic, sentence and sentence level helping the companies and political parties to understand the trending patterns and applications. In politics it helps in understand the perspective and reaction of people towards political parties, leaders and events. In this paper we have used twitter as one of the social media platforms to perform the sentimental analysis. Twitter is one of the widely used social media platform by various group of people to share their viewpoints. Everyday millions of views are shared on this platform which makes it suitable for performing the sentimental analysis. In the past there have been different machine learning models proposed but later various deep learning model has been introduced to increase the accuracy of sentimental classification.

## II. ALGORITHMS

Deep learning algorithms enhancements are a part of learning process in which they interact with outer data and find relevant errors moreover they themselves learn from the previous experience which contributes into their efficiency. Error detection and evaluation plays an important role in training and optimizing the data. This type of an approach helps devices to examine the error using minimum actions that eventually leads to improved efficiency in a particular context. The contradicting data should be eliminated as it leads to confusion and machines don't work at their normally efficiency.

Deep learning as well as Machine Learning counts in storage and processing of large data. Although computer does a good job in delivering more accurate and faster results to detect various opportunities and threats which are harmful, but more extra time and resources are also needed to train properly. The integration of ML with AI with support of technologies makes it more capable, reliable and accurate to produce large amount of data.

Support Vector Machine is a process of solving the Linear as well as Non-Linear Problems. It provides more accuracy when comparing it with Decision tree & LG. This can, able to work both on regression as well as classification problems. Outlier detection can also be done by using SVM. The Algorithm is slightly different as it maximizes the data points of all the classes and gives better results.

Logistic Regression is a process which is used to make predictions by using only few words. Like every analysis, it is used to describe the relationship B/W one Dependent variable with the other various levels of independent variables. A large portion of linear regression weights are also used to make predictions to use many words for this model.

Naïve Bayes is the process of Classification which bases on the theorem of Bayes. The Algorithm can be used for the classification on the training set which has More Dimensionality rate. Real Time Predictions can be done by using this Classifier. It can handle the discrete data as well as it is used to make the predictions based on the probabilities

Random Forest method overall increases the result and the linear model combinations increases. Adaptive Boosting is used to make a very strong classifiers by joining all weak one's Here the relationship between input and output layers is done by using the Multilayer Perceptron.

Neural Networks are just the layers which are same as neurons which are present in the human body & consists of layers of input as well as output. The weights in the network are nothing but the parameters which can learn and also uses many functions such as SoftMax and RELU. The quality of the output can be known by using the function named Loss

LSTM is a Special Type of Network precisely Recurrent Neural Network generally named RNN, It can vanishes the problem of gradient and the memory can also be retained for a Heavy period. It also consists of gates like input as well as

output and also uses the Sigmoid func. It also has the Information about the cell of previous by using the Keys.

### III. LITERATURE REVIEW

Ankita Sharma and Udayan Ghose has used Twitter dataset from twitter API to perform sentimental analysis to find out which candidate is trendier and more favoured in general election 2019 among the contesting candidates. Rapid miner's AYLIEN extension is used to extract named entities opinion polarity. Lexicon based approach for sentiment analysis and NRC dictionary-based approach is also used to perform the research. The result and conclusion obtain from the implementation was that candidate 1 is most liked in compared to candidate 2 and will be most likely to win the election [1].

According to Mohammad Karim Sohrabi and Fatemeh Hemmatia the best accuracy can achieve while using the SVM and ANN with the proposed preprocessing method. The study proposes an efficient preprocessing method for supervised sentimental analysis. Word2vec and TF-IDF method is used to convert the words array into numerical vectors. Python and RapidMiner was used for implementation to achieve the desired result [2].

As per Srishti Vashishtha and Seba Susan the model used i.e supervised fuzzy rule-based system for multimodal sentiment classification can classify the sentiments indicated in the video reviews on online media platform. (dataset was taken from CMU MOSI) unimodal and multimodal features are used for sentimental analysis. The Supervised SVM text accuracy was 71.21 and for Supervised SVM speech accuracy was 58.0. For Supervised text + speech accuracy was 57.5. Fuzzy Rule Based Text + Speech accuracy was 82.5. The Fuzzy rule based is better for multimodal sentiment analysis [3].

According to Sandeep Chaurasia, and Devesh Kumar deep learning techniques are efficient and effective for Sentiment and semantic analysis. Through Word2Vec and CNN the research is providing an accurate result for feature extraction technique for short sentence of movie review dataset. It cleanses the data properly and use CNN layer to extract good features for short sentence categorization and originates word vector from pre trained Word2Vec model. Deep convolution and Word2Vec neural network are used in research study for classification [4].

As per Priyanka Tyagi and Dr. R.C. Tripathi accurateness of sentiment classification is increased by applying the hybrid classification method of naïve bayes and KNN. To increase the accuracy of classifiers used for sentimental analysis using twitter dataset. N-gram algorithm is used for feature extraction. Naïve bayes classifier and SVM classifier is used. The dataset is taken from twitter API using Tweepy [5]

Saira Gillani and Sohail Asghar proposed that the effective method of finding the aspect based sentimental analysis. This article introduces a constructive method to examine the

sentiments by combining three operations – (a) Mining semantic feature (b) Transformation of extracted dataset using Word2vec. (c) Implementation of CNN for the opinion mining. The hyperparameters of CNN are tuned with GA [6].

Koyel Chakraborty, Surbhi Bhatia proposed the method for finding the false and negative tweets uploaded during the covid 19 pandemic. Fuzzy rule-based model was used in this paper. All the government should set up fact investigator in social and online media platform to prevent further sharing of misleading facts and information for cases which involve serious and severe concern [7].

As per Rui Ren and Dash Wu the model for forecasting stock market trends or movement direction. SVM model and logistics regression was used in this paper. Web crawler was used to collect the targeted textual document from internet i.e. news data and market data. The psychology of investors drives the stock market and the moderate return on Mondays is much low in comparison to the other weekdays [8]

The study conducted by Van Cuong Tran and Hyen Phan recommended a new method based on the feature ensemble model associated to twitter tweets containing fuzzy sentiments. The feature ensemble model was set up by combining the information from the 5 feature vectors obtained from lexical, word-type, semantic, sentiment polarity, and location of the words in twitter tweets consisting fuzzy sentiment phrases. Result from above phase is taken as input feature vector for CNN model which is used then used to arrange the sentiments of tweets into 5 categories. The feature ensemble model can effectively improve accuracy of a model for twitter tweets including fuzzy sentiment [9].

Sahar El\_Rahman, Feddah AlOtaibi and Wejdan AlShehri proposed a model for performing sentimental analysis of real data which was collected form twitter API. SVM, Maximum entropy, Decision tree, Random forest is used to classify the KFC and McDonald's data set, as per the study Maximum Entropy has provided the maximum result [10].

Manju Venugopalan and Deepa Gupta proposed the study of sentimental analysis on phone brands used over the past few years using the Naïve Bayes and SVM. As per their study SVM has provided the best accuracy [12].

Geetika Gautam and Divakar Yadav did the study of customer reviews. They proposed the different models of Machine learning analysing the product reviews. They have used Symantec analysis Naïve bayes and support vector machine. As per the study Symantec analysis provided better results [15].

Santosh Vishwakarma, Pragya Tripathi and Ajay Lala did a sentimental analysis on English tweets using Naïve Bayes and KNN algorithm on the rapid miner, in which they found out KNN is giving more accuracy [18].

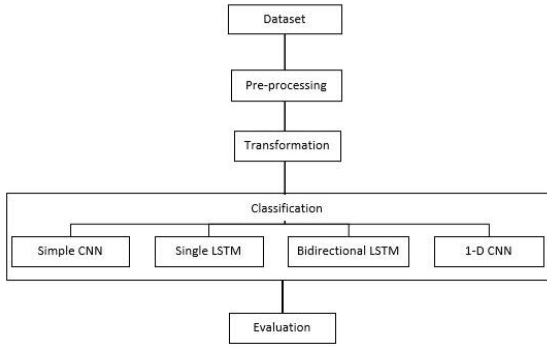


Fig. 1.

The above figure 1.1 describes various types of sentimental classification and the various Machine learning based approaches being used in various literature survey.

#### IV. PROPOSED MODEL

The proposed model development goes into the gathering of twitter records from the Kaggle. The twitter record contains following attributes: text ID, text and sentiment. Sentiment contains three classes or label which are neutral, positive and negative.

Further twitter records go into the pre-processing which includes following steps:

- Removing punctuation
- Removing stop words
- Lemmatization
- Tokenization
- Label encoding

In the next phase the data is transformed by sequencing and splitting.

##### A. LSTM - CNN Model

In the proposed model we have used CNN layer, LSTM layer and a dense layer.

For feature extracting we have used 1D CNN layer and the output of this layer is feed into the LSTM layer.

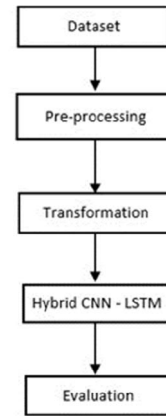


Fig. 2.

Figure 2 represents the proposed architecture.

```
model3.summary()
```

Model: "sequential_16"		
Layer (type)	Output Shape	Param #
embedding_16 (Embedding)	(None, 200, 40)	200000
conv1d_15 (Conv1D)	(None, 195, 20)	4820
dropout_16 (Dropout)	(None, 195, 20)	0
max_pooling1d_14 (MaxPooling)	(None, 39, 20)	0
lstm_14 (LSTM)	(None, 20)	3280
dropout_17 (Dropout)	(None, 20)	0
dense_14 (Dense)	(None, 3)	63
Total params: 208,163		
Trainable params: 208,163		
Non-trainable params: 0		

Fig 3

Above figure 3 represents the summary of the proposed model

#### V. RESULT ANALYSIS

The dataset is divided into training and testing where training accuracy and validation accuracy is calculated along with F1 score.

**Accuracy:** It is the proportion of no. of correct predictions to the total number of input samples. It can be calculated using below formulae.

$$Accuracy = \frac{\text{Number of Correct predictions}}{\text{Total number of predictions made}}$$

**F1 score:** It is done to measure a test accuracy. It is used to find out the mean between precision and recall. It mentions how accurately classifier has classified. High precision will give low recall, greater F1 score better the performance of model.

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}}$$

**Precision:** The total estimate max corrective of positive result divided by number of positives result figured by classifier.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

**Recall:** The number of rightly classified positives result divided by total samples.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

	Classifier	F1 Score	Recall_score	Precision_
0	SimpleRNN model	0.75	0.75	0.76
1	Single LSTM layer model	0.80	0.80	0.80
2	Bidirectional LTSM model	0.76	0.76	0.76
3	1D Convolutional model	0.81	0.81	0.82

Table 1

Table 1 represents the results and accuracy of the existing model.

	Classifier	F1 Score	Recall_score	Precision_
0	CNN-LSTM	0.84	0.84	0.84

Table 2

It represents the accuracy of the proposed model.

## VI. CONCLUSION

The model proposed in this paper gives the better accuracy in comparison to the other existing models. Therefore, sentimental analysis can be very useful for determining the views and emotions of the public about various trending topics (i.e. Business, products, politics). There are various machine learning models and deep learning models used to analyse and interpret the sentiments.

## VII. FUTURE WORK

In future we can also include emoticons or emojis for sentimental analysis and different languages other than English can also be used for classification.

## VIII. REFERENCES

1. Ankita Sharma and Udayan Ghose "Sentimental Analysis of Twitter Data with respect to General elections in India" in International Conference on Smart Sustainable Intelligent Computing and Applications under ICTETM2020.
2. Mohammad Karim Sohrabi and Fatemeh Hemmatian "An efficient preprocessing method for supervised sentiment analysis by converting sentences to numerical vectors: a twitter case study" in Springer Science+Business Media, LLC, part of Springer Nature 2019.
3. Srishiti Vashishtha and Seba Susan "Inferring Sentiments from Supervised Classification of Text and Speech cues using Fuzzy Rules" in International Conference on (ICCIDS 2019).
4. Amit Sharma, Sandeep Chaurasia and Devesh Srivastava "Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec" in ICCIDS 2019.
5. Priyanka Tyagi and Dr. R.C. Tripathi "A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data" in 2nd ICACSE-2019.
6. Adnan Ishaq and Saira Gillani "Aspect-based sentiment analysis using a hybridized approach based on CNN and GA" in 10.1109/ACCESS.2020.3011802, IEEE A
7. Koyel Chakraborty, Siddhartha Bhattacharya, Jan Platos, and Rajib Bag "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media" in journal homepage: www.elsevier.com/locate/asoc.
8. Rui Ren, Member, Desheng Dash Wu, Senior Member, and Tianxiang Liu "Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine" in IEEE Systems Journal.
9. Huyen Phan, Van Tran and Ngoc Nguyen, (Senior Member, Ieee) "Improving Performance of Sentiment Analysis of the Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model" In 10.1109/ACCESS.2019.2963702, IEEE Access.
10. Sahar El\_Rahman, Feddah AlOtaibi and Wejdan AlShehri "Sentiment Analysis of Twitter Data published in the International Conference on Computer & Information Science (ICCIS) in 2019 <https://doi.org/10.1109/ICCISci.2019.8716464> )

11. Nikhil Yadav, Omkar Kudale and Srishti Gupta "Twitter Sentiment Analysis Using Machine Learning for Product Evaluation" in Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT-2020)
12. Manju Venugopalan and Deepa Gupta "Exploring Sentiment Analysis on Twitter Data" <https://doi.ieeecomputersociety.org/10.1109/IC3.2015.7346686>
13. Radhi D. Desai "Sentiment Analysis of Twitter Data" in Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018).
14. Roshan Fernandes and Dr. Rio D'Souza "Analysis of Product Twitter Data though Opinion Mining" published in 2016 IEEE Annual India Coreference <https://doi.org/10.1109/INDICON.2016.7839025>
15. Geetika Gautam and Divakar Yadav "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis" in 2014 Seventh International Conference on Contemporary Computing (IC3) <https://doi.ieeecomputersociety.org/10.1109/IC3.2014.6897213>
16. Bac Le and Huy Nguyen "Twitter Sentiment Analysis Using Machine Learning Techniques" in Springer International Publishing Switzerland 2015 DOI: 10.1007/978-3-319-17996-4\_25
17. Yogesh Garg and Niladri Chatterjee "Sentiment Analysis of Twitter Feeds" in International Conference on Big Data Analytics <https://link.springer.com/conference/bigda> DOI: 10.1007/978-3-319-13820-6\_3
18. Pragya Tripathi, Santosh Vishwakarma and Ajay Lala "Sentiment Analysis of English Tweets Using RapidMiner" in 2015 International Conference on Computational Intelligence and Communication Networks. <https://doi.org/10.1109/CICN.2015.137>
19. Brandon Joyce and Jing Deng "Sentiment Analysis of Tweets for the 2016 US Presidential Election" in 2017 IEEE MIT Undergraduate Research Technology Conference (URTC) <https://doi.org/10.1109/URTC.2017.8284176>