

# Performance of Sentiment Analysis Approaches to Predict the Stock Markets

Seethiraju.L.V.V.D Sarma<sup>1</sup>, Dr Dorai Venkatasekhar<sup>2</sup>, Dr Gudipatu Murali<sup>3</sup>,

<sup>1</sup>Research Scholar, Dept of Information Technology, Annamalai University, Chennai, TN, [slvvdharma@gmail.com](mailto:slvvdharma@gmail.com).

<sup>2</sup>Professor, Dept of Information Technology, Annamalai University, Chennai, TN, [ramavenkatasekar@yahoo.co.in](mailto:ramavenkatasekar@yahoo.co.in).

<sup>3</sup>Professor, Dept of Computer Science & Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, AP.

**Received** 2022 March 25; **Revised** 2022 April 28; **Accepted** 2022 May 15.

**Abstract:** Stock markets are very important for every country and company to increase the economy of the country. Many researchers are trying to develop the accurate prediction of analysis of stock markets. Sentiment Analysis (SA) is the sub-domain in text mining. For the prediction of various opinions and text messages in several social networking sites (SNS). The role of sentiment analysis in stock markets plays a significant role in predicting accurate results in stock market analysis. Many companies are listed in the stock markets to increase the business and improve the stock prices of the specific company. For the past many years many people are investing in the stock markets. In this paper, the performance of various sentiment analyses in machine learning (ML) and Deep Learning (DL) approaches are discussed and analyzed the performance of the algorithms based on the prediction. The parameters such as Sensitivity, Specificity, Accuracy, F-measure and Area under curve (AUC).

**Keywords:** Sentiment Analysis (SA), Social Networking Sites (SNS), Machine Learning (ML), Deep Learning (DL), Stock Markets.

## Introduction

"Sentiment Analysis (SA)" is the word firstly used by [1], this gains the huge attraction in the research world. SA is most widely used to analyze sentiments based on opinions and text analysis. This can be applied to many applications to get the opinions of users or clients. Sentiments can be analyzed with the text a message belongs to the various applications. These text messages are considered as feedbacks given several applications. Many websites and mobile apps providing the stock market analysis based on the feedback of the users given to the companies. Due to the random walk behavior of the stock markets, it is very difficult to predict the stock prices in time series. To gain from these stock prices, people are searching for more and more tools and software's to reduce the risk prediction. Sentiment Analysis deals with several real-time data belongs to social networks sites (SNS) such as face book, twitter etc. In stock prices prediction, the time-series prediction is a very difficult type based on the issues present in predictive modeling. On the other side, SA is applied on several fields such as reviews of movies [2], behavior of financial products based on sales [3], to improve the services and items that improves the services [4], by analyzing the customer reviews the services are improved.

SA uses Natural Language Processing (NLP) for improving the performance of various machine learning algorithms. These applications develop the models for classification of tweets as positive (POS), negative (NEG) and neutral (NEU) sentiments [5], by analyzing several opinions about the items that are available in blogs, reviews, and comments [6]. Stock market prediction is a very difficult task to analyze accurate stock prices. The stock prices are predicted and analyzed by the movements of the prices that are widely used in numerous fields such as finance, trading, statistics, and system analysis. Professionals in the stock markets use technical analysis to analyze the stock market items/indices to take accurate decisions for investment. In this paper, numerous algorithms are discussed and analyze the performance of ML and DL algorithms to predict the daily stock market prices. Figure 1 explains about the stages of stock market predictions using ML and DL algorithms.



Figure 1: Architecture Diagram

### Role of Sentiment Analysis in Stock Markets

Among all the available domains, stock market prediction is very difficult task. SA is totally based on the positives (POS), Negatives (NEG) and neutral (NUE). Based on the polarity score the sentiment analysis is calculated by using the equation (1).

$$\frac{\sum \text{POS}_a - \sum \text{NEG}_a}{\sum t_a} \dots \dots \dots \text{Equation} - (1)$$

Where  $\text{POS}_a$  represents the  $a^{\text{th}}$  positive tweet and  $\text{NEG}_a$  represents the  $a^{\text{th}}$  negative tweet and

$t_a$  represents the total tweets

### Role of Deep Learning in Stock Market Sentiment Analysis

DL plays the significant role in stock market analysis using sentiment analysis. DL consists of multiple layers that can work very efficient on stock market prediction. DL plays the significant role in stock market analysis using sentiment analysis. DL consists of multiple layers that can work very efficient on stock market prediction. The input layer, hidden layer and output layers are present in the DL models. Based on the input dataset, the hidden layer consists of 'N' number of neurons to identify the relationships among the stock data, data collected from various sources. Figure 2 explains about the CNN architecture applied on stock market data predictive analysis.

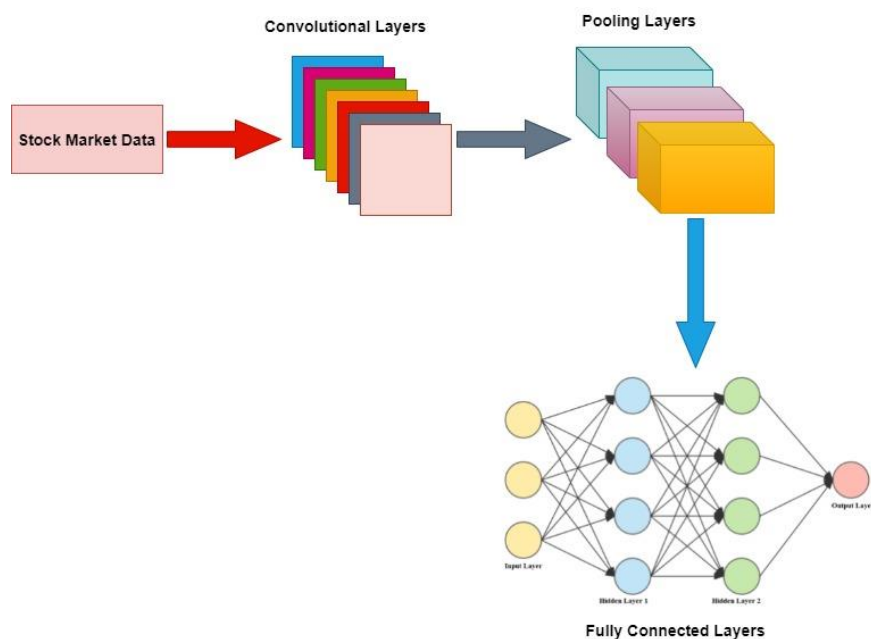


Figure 2 Deep Learning (DL) for Stock Market Analysis

### **Literature Survey**

Pagolu et al., [7] proposed the new model that utilized the two distinctive literary portrayals, Word2vec and N-gram, for dissecting the public feelings in tweets. For two distinct printed portrayals, Word2vec and N-gram, for examining the public feelings in tweets. Opinion investigation and managed AI standards to the tweets extricated from Twitter and dissect the relationship between's financial exchange developments of an organization and feelings in tweets. Extravagantly, positive news and tweets in web-based media about an organization would urge individuals to put resources into the supplies of that organization, and accordingly, the stock cost of that organization would increment. The author showed that a solid connection exists between the ascent and falls in stock costs with the public opinions in tweets.

Mohan et al., [8] proposed the new model which shows the improved accuracy of the stock prices prediction using a huge amount of time series information that belongs to the news articles, by using DL models. The dataset is collected from several sources that consists of stock prices belongs to daily basis for top 500 companies up to 5 years, and also this contains 264,500 news articles of various companies belongs to financial status. For the processing of large-size datasets, cloud computing is used for the training of predicting models and improves the performance in given stock in real-time.

Sharma et al., [9] introduced the effective approach that predicts the prices of stock markets based on the previous date database. Pinto et al., [10] this review give a methodical audit 57 chose papers utilizing time-series, message mining, and feeling examination applied to anticipate monetary financial exchange conduct. Through this exploration, it was seen that the utilization of information from online media and web destinations is a compound wellspring of data, giving a superior forecast. Be that as it may, the choice and mix of this information in a pertinent manner are still impediments found in the proposed models.

Khatri et al., [11] proposed the new SA to classify the social media data using classification algorithm. This algorithm is used to classify the mood of the users by using comments in social media. The comments are classified as various emotions. The pre-trained model is used to predict the vector of Bombay Stock Exchange index value for  $(n+1)^{th}$  days.

Zad et al., [12] break down the audits of the strategies of message-based feeling examination pipeline including preprocessing, angle extraction, highlight choice, and characterization methods utilized by researchers as of late. It likewise studies various utilization's of semantic investigation with regards to web-based media, advertising, and item surveys. Wilson et al., [13] proposed the new model that shows the differences between layers in the documents.

In sentiment layer, the classification of sentiments is done with the specific aspects that define the starting of this step. Current ways to deal with SA are fundamentally founded on administered realizing which depends on manual marked examples like business item audits in Google, Amazon, and other informal organizations item surveys in which the general demeanor of clients is expressly indicated in the audit utilizing stars, numbers, and so on Likewise move learning models [14] with AI models [15], can further develop text-based applications.

Su et al., [16] developed and introduced the new learning model that translates the words to extract the complete and useful meaning based on the emotions given in the form of text. This method uses the interconnectivity that represents the construction of paragraph or phrase. **Performance Analysis**

- The performance is analysed by using the confusion matrix/table to show the performance of proposed models.
- It is also called as error matrix.
- This matrix shows the overall predictions that are correct or incorrect.
- Below table shows the matrix parameters:

|                    | Actual Positive | Actual Negative |
|--------------------|-----------------|-----------------|
| Predicted Positive | True Positive   | False Positive  |
| Predicted Negative | False Negative  | True Negative   |

**Precision**

This is one of the parameter shows the original positives that are correctly predicted.

$$\text{Precision} = \frac{\text{No. of TP}}{\text{No. of TP} + \text{No. of FN}}$$

**F1 Measure**

This is the measure that shows classification of prediction values as ‘positive’ or ‘negative’.

$$\text{F1 Measure} = 2 \times \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

**Accuracy:** This will calculate the overall accuracy of the result.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

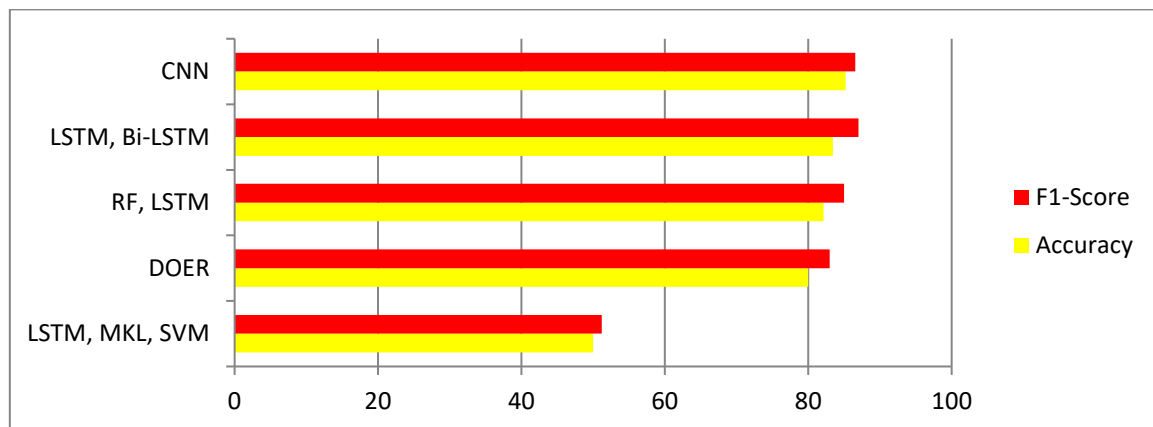
**Recall:** Appropriate when minimizing false negatives is the focus.

$$\text{Recall} = \frac{\text{TP}}{\text{No. of TP} + \text{No. of FN}}$$

**Comparison between various Sentiment Analyses in Several Domains**

| Author               | Algorithm      | Dataset                                         | Measures       |
|----------------------|----------------|-------------------------------------------------|----------------|
| Li et al., [17]      | LSTM, MKL, SVM | Stock Exchange data belongs to Hong Kong (HKEx) | 50%-Accuracy   |
| Luo et al., [18]     | DOER           | SemEval (2014, 2015, 2016)                      | 83%- F1 Score  |
| Kraus et al., [19]   | RF, LSTM       | Amazon dataset                                  | 85%- F1 Score  |
| Wu et al., [20]      | LSTM, Bi-LSTM  | Chinese VA, Facebook, Emobank                   | 86.56-F1 Score |
| Alharbi et al., [21] | CNN            | SemEval-2016                                    | 87%- F1 Score  |

**Table 1: Comparison between various Algorithms**



**Figure 2: Performance Analysis**

## Conclusion

In this paper, the role of sentimental analysis in predicting the stock markets is discussed clearly. Sentiment analysis is part of text mining and machine learning. SA can apply to many domains and also develop many applications. Prediction plays a significant role in stock market prediction which can face more challenges using ML algorithms. This paper mainly focused on discussing the various models that are applied to the prediction of stock market analysis using numerous ML algorithms. Several approaches which are basic methods achieved the medium accuracy of stock prediction. An ensemble and intelligent approaches are to be developed to overcome the accuracy issues.

## References

- [1] G. Gidofalvi and C. Elkan, "Using News Articles to Predict Stock Price Movements", University of California, San Diego, Jun 15, 2001
- [2] Y. Kara, M. A. Boyacioglu and O. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange", Published in Expert Systems with Applications, vol. 38(5), May1, 2011.
- [3] K. Kim, "Financial time series forecasting using support vector machines", Published in Neurocomputing 2003, DOI: 10.1016/S0925-2312(03)00372-2.
- [4] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques", Expert Systems with Applications: An International Journal, vol. 42, no. 1, January 2015, pp. 259-268.
- [5] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Sentiment Analysis of Twitter Data", LSM '11 Proceedings of the Workshop on Languages in Social Media pp. 30-38.
- [6] G. Vinodhini and R.M. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, no. 6, June 2012.
- [7] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016, pp. 1345-1350, doi: 10.1109/SCOPES.2016.7955659.
- [8] S. Mohan, S. Mullanpudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis," 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), 2019, pp. 205-208, doi: 10.1109/BigDataService.2019.00035.

- [9] A. Sharma, D. Bhuriya and U. Singh, "Survey of stock market prediction using machine learning approach," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 506-509, doi: 10.1109/ICECA.2017.8212715.
- [10] N. Pinto, L. da Silva Figueiredo and A. C. Garcia, "Automatic Prediction of Stock Market Behavior Based on Time Series, Text Mining and Sentiment Analysis: A Systematic Review," 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2021, pp. 1203-1208, doi: 10.1109/CSCWD49262.2021.9437732.
- [11] S. K. Khatri, H. Singhal and P. Johri, "Sentiment analysis to predict Bombay stock exchange using artificial neural network," Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization, 2014, pp. 1-5, doi: 10.1109/ICRITO.2014.7014714.
- [12] Zad, Samira & Heidari, Maryam & Jones, James & Uzuner, Ozlem. (2021). A Survey on Concept-Level Sentiment Analysis Techniques of Textual Data. 0285-0291. 10.1109/AIIoT52608.2021.9454169.
- [13] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of human language technology conference and conference on empirical methods in natural language processing, pp. 347–354, 2005.
- [14] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in IEEE 2020, International Symposium on Technology and Society (ISTAS20), ISTAS20 2020, 2020.
- [15] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021.
- [16] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, "Hidden sentiment association in chinese web opinion mining," in Proceedings of the 17th international conference on World Wide Web, pp. 959–968, 2008.
- [17] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong," Information Processing & Management, vol. 57, no. 5, p. 102212, 2020.
- [18] H. Luo, T. Li, B. Liu, and J. Zhang, "Doer: dual crossshared rnn for aspect term-polarity co-extraction," arXiv preprint arXiv:1906.01794, 2019.
- [19] M. Kraus and S. Feuerriegel, "Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees," Expert Systems with Applications, vol. 118, pp. 65–79, 2019.
- [20] C. Wu, F. Wu, S. Wu, Z. Yuan, J. Liu, and Y. Huang, "Semisupervised dimensional sentiment analysis with variational autoencoder," Knowledge-Based Systems, vol. 165, pp. 30–39, 2019.
- [21] A. S. M. Alharbi and E. de Doncker, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information," Cognitive Systems Research, vol. 54, pp. 50–61, 2019.

#### **About Authors**



Dr D.Venkatesekhar has Received the B-Tech Degree In Electronics & Communication Engineering From JNTU COLLEGE OF ENGINEERING Kakinada in 1987, MS (Engg) by Research In Information Technology from Annamalai University in 2002, and PhD In Computer Science and engineering in 2015 from Annamalai University. He has Teaching Experience of 30+ Years. His Field of Interest Include Among Others Big Data, Data Analytics, Data Mining, Image Processing. Computer Networks And Data Communications. He is currently guiding 8 PhD Scholars.



Dr. Gudipati Murali has completed his Degree from Nagarjuna University, in 1996. Master's degree in Computer Science and Engineering from JNRNU, Rajasthan in 2005 and Ph.D in Computer Science and Engineering in 2013 from ANU, Guntur. Currently, Working as a professor in Computer Science and Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, AP. Has teaching experience of 20 years and published several papers in various national and international journals. Currently nine members of scholars are working under his guidance. Four of his scholars has been awarded Ph.D degree from Acharya Nagarjuna University. His field of interest includes Computer Networks, Data Mining, Big Data, Data Analytics, Digital Streaming, Image Processing, VANET's and Internet of Things.

I am Seethiraju.L.V.V.D Sarma, Research Scholar in Annamalai University. I am doing my research on Sentiment Analysis in Stock Markets.

