IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stock Market Analysis Using Machine Learning And Sentiment Analysis

[1]Gautami Bafna, [2]Manish Patil, [3]Poojan Modi,[4]Rutika Shinde,[5]Ashish Awate

[1]Final Year Student , [2]Final Year Student, [3]Final Year Student,[4]Final Year Student,[5]Assistant Professor
Computer Engineering Department,
Shri Vile Parle Kelavani Mandal's,Institute Of Technology,
Dhule,India

Abstract: Stock market prediction has always been a pivotal and critical task since years, due to vigorous fluctuations in stock values. It helps in determining future stock value for investors, buyers and sellers in the commercial market. Looking towards today's world economy, prediction of the stock market has become a much more important factor .More accurate and efficient predictions may yield significant profits, stabilize the world's financial condition to some extent. For this we need to build a model for the basic stock market on the basis of machine learning algorithms and sentiment analysis. We have surveyed eight papers and studied them thoroughly and understood that stock market values can be predicted upto great extent using machine learning by its various algorithms and through sentiment analysis we may get a clear idea about whether stock market prices will be high or low. So, in this paper we have used certain machine learning algorithms and sentiment analysis on data retrieved from social media, financial news, historical stock prices and blogs to predict stock market values.

Index Terms - machine *learning, sentiment analysis, random forest, support vector machine(svm), artificial neural network(ann)*

I. INTRODUCTION

In this paper we will build a stock prediction model using machine learning and sentiment analysis. Now let's understand what actually the **machine learning** and **sentiment analysis** is:

- **Machine Learning:**

  Basically machine learning means a type of artificial intelligence which learns itself from the data provided and then works on it and applies that learning on a system without any need for humans.[10]

  The Machine Learning Algorithms are classified as:

1. *Supervised Learning:*

In supervised learning, the machine is trained using labeled data .Then the machine is compared with new data , so that supervised learning algorithm analyzes the training data and thus produces the outcome through labeled data.[11]

2. *Unsupervised Learning:*

In unsupervised learning, the machine is not trained using any sort of labeled data. Instead, it sorts the data according to their similarities and patterns and groups them accordingly. It doesn't give any correct output but gives inferences from datasets.[11]

3. *Semi-Supervised Learning:*

In semi-supervised learning, a little amount of labeled data is used and a large amount of unlabelled data is used .It lies between supervised and unsupervised learning. The machines using these algorithms actually improve the learning accuracy. [11]

4. *Reinforcement Learning:*

In reinforcement learning, the errors or rewards are discovered ,actions are produced by interacting with the environment. This algorithms allows the machines to automatically determine the ideal behavior of the system within a particular context to optimize the performance of machines.[11]

- **Sentiment Analysis:**

Sentiment analysis means to process the data and then categorize that data to positive, negative or neutral. Sentiment analysis is also referred to as opinion mining .So, it's a machine learning tool which is used to analyze texts, punctuations, spaces and more such noises to detect polarity, if positive or negative. Its algorithms are Naive Bayes , Linear Regression, etc. [12]

So, in this paper we will use supervised machine learning algorithms that are artificial neural networks (ann.), support vector machines (svm), random forest and through sentiment analysis we will check whether the sentiment of our data is positive, negative or neutral. If sentiment of model reflects positive angle it means stock values are high and there are higher chances of profit and if sentiment reflects negative angle it means stock values are low and an individual may suffer losses in the market.

II. LITERATURE SURVEY

Before getting into actual project we first surveyed and found around eight stock market prediction related research papers and studied those papers thoroughly and then we decided that from where and what type of data need to be collected and processed, on which machine learning algorithms we should work and how to apply sentiment analysis accordingly and those papers we have briefly stated in our paper.

Those papers are:

A. Stock market prediction using machine learning classifiers and social media, news:

1. Theme: To study and analyze the impact of financial news data and social data on stock market prediction accuracy for ten subsequent days using certain algorithms. [1]

2. Proposed Method: They proposed a framework for stock market forecasting using financial news and social media. The proposed system relies on each aspect of the data and the system itself for achieving accurate predictions. So, the proposed system for stock prediction is divided into eight subsystems. [1]

3. Experiment & Result: The experimental results show that using social media and financial news, highest prediction accuracies of 80.53% and 75.16% are achieved respectively. Also the paper shows that New York and Red Hat stock markets are difficult to predict, New York and IBM stocks are more affected by social media, whereas in London and Microsoft stocks by financial news. Random forest classifier is found to be more efficient and highest accuracy of 83.22% is achieved.[1]

B. Stock Market Prediction Using Machine Learning Algorithms :

1. Theme: To find the best model to predict the values of the stock market. In this paper they are going to present a more feasible method to predict the stock market fluctuations with higher accuracy. [2]

2. Proposed Method: This paper focuses on predicting the stock value using random forest and SVM.

3. Experiment & Result: Results show that on collecting raw data apply data mining techniques and random forest algorithm to process data and accuracy is found to be 0.808.

C. Research on Stock Price Prediction Method Based on Convolutional Neural Network :

1.Theme: To build a model CNN (Convolution Neural Network)-based in-depth learning method for predicting stock prices.

2. Proposed Model: This paper proposes a stock price prediction model based on convolution neural network,which has automated-learning ability. After combining the characteristics features of CNN and Thai stock market, the data set is trained and tested after pretreatment. [3]

3. Experiment & Result: On that basis, three stocks namely BBL, CAPLL and PTT are listed on the Thai Stock Exchange and are tested and compared with the actual stock price. The results show that the model based on CNN can effectively identify the variations in stock prices and predict it which can provide valuable reference for stock price prediction. The Paper says that models with high accuracy can satisfactorily help in the financial world. [3]

D.Study on the prediction of stock price based on the associated network model of LSTM:

1. Theme:To design a model using deep network parallel predict the opening price, the lowest price and the highest price of a stock on the next day according to the historical stock prices and other data constraints.[4] Proposed Model: They have proposed an associated deep recurrent neural network model by providing multiple inputs and multiple outputs based on a long short-term memory network. The associated network model has the capability to predict the opening price, the lowest price and the highest price of a stock parallel. The associated network model was compared with the LSTM network model and deep recurrent neural network model. The feasibility and accuracy of the Associated Net are verified by comparing the model with the LSTM network model and the LSTM deep recurrent neural network model.[4]

2. Experiment & Result: Experiments show that the average accuracy of the Associated Net model found to be better than the remaining two models and it can also predict multiple values parallel. The average accuracy of each predicted value is over 95%. [4]

E.An innovative neural network approach for stock market prediction:

1. Theme: This paper aims to build an innovative neural network model to get better stock market predictions.

2. Proposed Model: They proposed the deep long short-term memory neural network (LSTM) with embedded layer and the long short-term memory neural network with automatic encoder to predict the stock market. In these two models, they use the embedded layer and the automatic encoder, respectively, to vectorize the data, in a bid to forecast the stock through a long short-term memory neural network. [5]

3. Experiment & Result: The experimental results show that the deep LSTM with embedded layer is better. Speci□cally, the accuracy of two models is 57.2 and 56.9%, respectively, for the Shanghai A-shares composite index. Furthermore, they are 52.4 and 52.5%, respectively, for individual stocks. They demonstrate research contributions in IMMT for neural networks - based on □nancial analysis. [5]

F. Survey on Combined Swarm Intelligence and ANN for Optimized Daily Stock Market Price:

1. Theme: This paper makes a survey of the use of Swarm Intelligence in a stock market application. The paper initially describes the details of a stock market, SI and its various types of algorithms and finally shows some recent SI algorithm based approaches that can be used for stock market prediction. [6]

2. Proposed Model: To improve the efficiency of SI and make optimized results, SI is combined with other approaches like Artificial Neural Network (ANN), Machine Learning ML etc.

3. Experiment & Result: They found that by combining SI and ANN the model produces more accurate and optimized results for stock price prediction than the combination of SI and machine learning. [6]

G. Developing a Prediction Model for Stock Analysis:

1. Theme: To use Neural Networks to predict the best stock price. In this work, various prediction algorithms are studied and analyzed to build a prediction model.

2. Proposed Model: This paper proposes the uses of linear regression and neural networks. To build a stock prediction model. The prediction model will be based on monthly prediction and daily prediction to forecast the next day market price. This model predicts the open value of the next day in the market.

3. Experiment & Result: A comparative study of the three algorithms that are: Multiple Linear Regression, Support Vector Machine and Artificial Neural Network are done and found that the stock price is predicted by sentiment analysis with the best forecasting algorithm. [7].

H. Stock Trend Prediction Using News Sentiment Analysis:

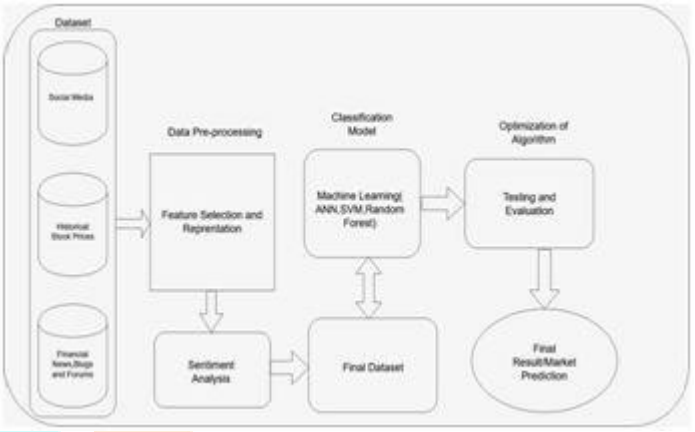
1. Theme: This paper is about taking non quantifiable data such as financial news articles of a company and predicting its future stock trend with the help of news sentiment analysis. They have assumed that news articles have an impact on the stock market, so they have attempted to study the relationship between news and stock trends. [8]

2. Proposed Model: They have created three different classification models which depict polarity of news articles to check whether positive or negative and to build a model they use SVM and naive bayes as testing classifiers.

III. Experiment & Result: Experiments are conducted to evaluate various aspects of the proposed model and efficient results are obtained in all of the experiments. The accuracy of the prediction model is more than 80% and in comparison with news random labeling with 50% of accuracy; the model has increased the accuracy by 30%.[8]

IV. PROPOSED SYSTEM

After surveying all the papers we have found that through news articles, historical data we can build efficient and accurate model using machine learning and sentiment analysis. So, in this project we will take datasets like news from social media, historical stock prices, financial news, blogs and forums. The data taken from those resources will be pre-processed using feature selection and representation methods through which we can do sentiment analysis and we will get a final dataset to build our model .The final dataset which we will receive from sentimental analysis will be passed to three classification algorithms of machine learning. Probably, we are going to use three classification algorithms namely Artificial Neural Network, Support Vector Machine and Random Forest. Three graphs will be plotted accordingly and will test the data to check optimized algorithms and then it will be evaluated. Hence, we will get the final result and present it in an app to make it easy for investors, sellers and buyers to view stock market trends. Here is the design of our proposed system:



fiure.1.proposed model

Now, we will study about the proposed system step by step in brief:

A. Data Collection:

Dataset to build prediction models will be collected from Social media, historical stock prices, financial news,blogs and forums. Factors to be considered are opening value, closing value, etc.

B. Data Pre-processing:

All the data collected need to be pre-processed for the classification because the data will not be structured and might contain noise, unnecessary information .So, for this we need to filter and structure the data.

C. Sentiment Analysis:

The sentiment of data is to be checked in this phase for which we will use Monkey Learn R package. Pre-trained models for opinion mining are much more efficient because we can do data analysis correctly. But, if we want maximum accuracy, we should train a custom model which uses our own data and constraints for sentiment analysis. By doing so we can enable the machine learning model to learn from industry-specific expressions but also understand our criteria and tell us what to consider positive or negative. We will use MonkeyLearn to create a custom model for sentiment analysis.[9]After the proper sentiment detection is done we will get our final dataset through which classifying models will be built.

D. Classification Model:

After thoroughly studying the research papers we have analyzed that most of the papers strongly favours SVM,Random Forest and ANN.So, we will use these three classifying machine learning algorithms to build our model.

E. Testing And Evaluation:

By applying those three classification algorithms we get graphs for each algorithm from which we will get a sum of squared error for training data, testing data and cross validation and we will get optimized and best from the three algorithms. i.e., we actually test the classifiers and finally evaluate the optimized algorithm among the three.

F. Result:

Finally, we will get a stock prediction model which we will represent in form of an app which will make investors, sellers and buyers to check stock market trends.

V. CONCLUSION

In this project, we have proposed a design of a stock predicting model which will be presented in the form of an android app. The predicting model is based on different Machine Learning algorithms and Sentiment Analysis, which depicts different trends indicating future stock prices of the company. This app is designed and supposed to help the investors, traders to invest their money for good profits. Which will decrease the chances of losing their money in the stock market also it will save time for the investors for choosing the proper stocks.

Authors and Affiliations:

1.Gautami Bafna

Final Year Student of Computer Engineering at Shri Vile Parle Kelavani Mandal's Institute Of Technology ,Dhule ,Maharashtra , India.

2.Manish Patil

Final Year Student of Computer Engineering at Shri Vile Parle Kelavani Mandal's Institute Of Technology, Dhule, Maharashtra , India.

3.Poojan Modi

Final Year Student of Computer Engineering at Shri Vile Parle Kelavani Mandal's Institute Of Technology, Dhule, Maharashtra , India.

4.Rutika Shinde

Final Year Student of Computer Engineering at Shri Vile Parle Kelavani Mandal's Institute Of Technology, Dhule, Maharashtra , India.

5.Ass.Prof.Ashish Awate

Currently working as Assistant Professor in Department of Computer Engineering at Shri Vile Parle Kelavani Mandal's Institute Of Technology, Dhule, Maharashtra , India.

REFERENCES

[1]Wasiat Khan, Muhammad Awais Azam,Mustansar Ali Ghazanfar, Ahmed S. Alfakeeh, Amin Karami, Khaled H. Alyoubi,"Stock market prediction using machine learning classifiers and social media, news", In Springer-Verlag GmbH Germany, part of Springer Nature, 2020.

[2]K. Hiba Sadia, Aditya Sharma, SarmisthaPadhi, Adarrsh Paul, Saurav Sanyal,"Stock Market Prediction Using Machine Learning Algorithms",In International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April 2019.

[3] Sayavong Lounnapha ,Wu Zhongdong ,Chalita Sookasame,''Research on Stock Price Prediction Method Based on Convolutional Neural Network",In International Conference on Virtual Reality and Intelligent Systems (ICRIS),2019.

[4] Guangyu Ding1, Liangxi Qin1,"Study on the prediction of stock price based on the associated network model of LSTM",In International Journal of Machine Learning and Cybernetics ,2019.

[5] Xiongwen Pang, Yanqiang Zhou1,Pan Wang1,Weiwei Lin2, Victor Chang,"An innovative neural network approach for stock market prediction",In Springer Science+Business Media, LLC, part of Springer Nature 2018.

[6] Mr. Pankaj K. Bharne and Dr. Sameer S. Prabhune, "Survey on Combined Swarm Intelligence and ANN for Optimized Daily Stock Market Price",2017.

[7]R.Yamini Nivetha,Dr. C. Dhaya,"Developing a Prediction Model for Stock Analysis",DOI 10.1109/ICTACC.2017.11IEEE,In International Conference on Technical Advancements in Computers and Communications,2017.

[8] Kalyani Joshi ,Prof. Bharathi H. N, Prof. Jyothi Rao,"Stock Trend Prediction Using News Sentiment Analysis",DOI:10.5121/ijcsit.2016.8306,In International Journal of Computer Science & Information Technology (IJCSIT) Vol 8, No 3, June 2016

[9] https://monkeylearn.com/blog/how-to-do-sentiment-analysis-in-r/

[10] https://www.investopedia.com/terms/m/machine-learning.asp

[11] https://expertsystem.com/machine-learning-definition/

[12] https://monkeylearn.com/blog/sentiment-analysis-machine-learning/