



SURVEY PAPER ON STOCK MARKET PREDICTION

K. A. Jamadar¹, S. R. Khan², N. M. Kharade³, N. S. Bargir⁴, Rohit. D. Mane⁵

^{1,2,3,4,5} Department of Computer science and Engineering, Dr.J.J.Magdum college of Engineering, Jaysingpur, India

Email ID: kayyunjamadar2@gmail.com, sr99khan@gmail.com, nitakharade00@gmail.com, namirabargir20@gmail.com, rohit.mane@ijmcoe.ac.in

ABSTRACT

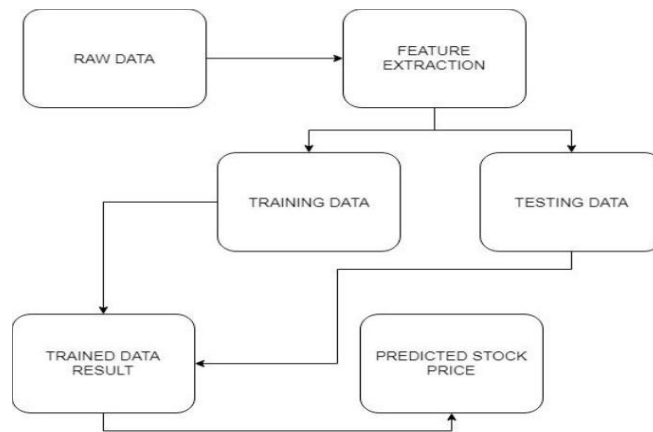
Stock market prediction and analysis are some of the complex task to perform . There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision. However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data.

Keywords: *Machine Learning, SVM, Random Forest Algorithm, Yahoo Finance, Dataset, Stock, Stock Market.*

1. INTRODUCTION

In the global financial market trading is one of the most important tasks. The stock market is basically a collection of various buyers and sellers of the stock. A stock market prediction is an act of trying to determine the future value of another financial instrument traded on a currency exchange [1]. This paper describes stock forecasts using Machine Learning. Structured language is used to predict the stock market using machine learning by Python. In this paper, we propose a Machine Learning (ML) method that will be trained from available stock data and gain intelligence and apply the knowledge gained from accurate predictions. In this context this study uses a Random Forest algorithm and Support Vector Machine (SVM) method to forecast the prices of capital and small capital in three different markets, using daily and daily price levels. Machine learning involves artificial intelligence which empowers the system to learn and improve from past experiences without being programmed time and again [6]. The prediction is expected to be robust, accurate, and efficient. In the system, we firstly collect raw data from the yahoo finance dataset. It is the initial step. It generally deals with the collection of the right dataset. The data mainly consists of the previous year's stock prices. Yahoo Finance has historical data which contains various companies' stock data that tells the opening and closing price of the market ranging from the highest and lowest point. It has data for many years, which is divided into 3 parts ----1. Daily updates 2. Weekly updates, 3. Monthly updates. The datasets of the stock market prediction model include details like the closing price opening price, the data, and various other variables that are needed to predict the object variable which is the price in a given day. The previous models used traditional methods of prediction like multivariate analysis with a prediction time series model. Stock market prediction outperforms when it is treated as a regression problem but performs well when treated as a classification [7]. The aim is to design a model that finds high accuracy results. In this system, we take SVM and the random forest algorithm to predict stock. Predictive methods like the Random forest technique are used for the same. The random forest algorithm follows an ensemble learning strategy for classification and regression. The random forest takes the average of the various subsamples of the dataset, this increases the predictive accuracy and reduces the over-fitting of the dataset [3]. The Support Vector Machine (SVM) can be used for both classification and regression. It has been observed that SVMs are more used in classification-based problems [3][4] . In the SVM technique, we plot every single data component as a point in n-dimensional space (where n is the number of features of the dataset available) with the value of feature being the value of a particular coordinate, hence classification is performed by finding the hyperplane that differentiates the two classes explicitly[2][3]. In both methods, we divide the data into a ratio of 80:20 where 80% is for the training set and the rest of 20% data is for the testing data set [5]. We take both random forest algorithm and SVM algorithms results and we find the average of both results for high accuracy prediction results.

ML ARCHITECTURE



Modules:

1. Data Collection
2. Pre – Processing
3. Training the machine using random forest
4. Training the machine using support vector machine
5. Calculating the average of both the results obtained by the algorithms

2. LITERATURE REVIEW

1. DivyaYadav¹, Dr. N.Thillaiarasu² (2020) “Stock Market Prediction Using Machine Learning (ML) Algorithms.” School of Computing Science & Engineering, Galgotias University, Greater Noida, U.P

In the global financial market trading is one of the most important tasks. A stock market forecast is an act of trying to determine the future value of another financial instrument traded on currency exchange. This paper describes stock forecasts using Machine Learning. In this context, this study uses a Random Forest and Vector Machines method to forecast the prices. Random Forest algorithm was found to be the most frequently used technique for stock price prediction due to its good performance and accuracy. Python is more suitable for machine learning, artificial intelligence. AI developers prefer Python over Java because of its ease of use, accessibility, and simplicity. Java has a better performance than Python but Python requires lesser code and can compile even when there are bugs in your code. Python as a language has a great community behind it. Any problems encountered can be easily resolved with a trip to Stack Overflow. Python has many powerful tools ready for scientific computing. Packages such as Numpy, Pandas, and SciPy are freely available and well documented. Packages like these can be surprisingly reduced, and simplify the code needed to write a given program. This makes iteration faster.

2. Wasiat Khan Mustansar Ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled H. Alyoubi, Ahmed S. Alfakeeh (2020), “Stock market prediction using machine learning classifiers and social media, news”.

Drawback is the previous results indicate that the stock price is unpredictable when the traditional classifier is used. We got to know in Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values. Factors considered are open, close, low, high and volume.

3. ALGORITHM’S

a) Random Forest Algorithm:

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression [1] . One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification [5]. It performs better results for classification problems.

b) Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems [1]. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future[8]. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane[8]. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine[8].

4. MODULES

1) Data Collection:

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external. Here the data mainly consists of the previous year stock prices. Initially, we will be analyzing the Yahoo finance dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately. The data set will be stored in the CSV file (Comma Separated Value).

1) Data Pre-processing:

Data pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

2) Predicting the values using Random Forest Algorithm

Training the machine is similar to feeding the data to the algorithm to touch up the test data. The training sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data[1]. Tuning models are meant to specifically tune the hyperparameters like the number of trees in a random forest[2]. Finally, we will calculate a cross-validated score, for individual sets of hyperparameters. Then, we select the best hyperparameters. The idea behind the training of the model is that we some initial values with the dataset and then optimize the parameters which we want to in the model. This is kept on repetition until we get the optimal values. Thus, we take the predictions from the trained model on the inputs from the test dataset. Hence, it is divided in the ratio of 80:20 where 80% is for the training set and the rest 20% for a testing set of the data.

3) Predicting the values using Support Vector Machine Algorithm

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen[8]. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. Hyperplanes are decision boundaries that help classify the data points[8]. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane[8]. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

4) Calculating the average Result

This Module helps to calculate the average obtained by Random Forest and Support Vector Machine. The average of efficiency of both the algorithms will be calculated to learn about its coherency.

5. RESULT

For Random Forest Algorithm

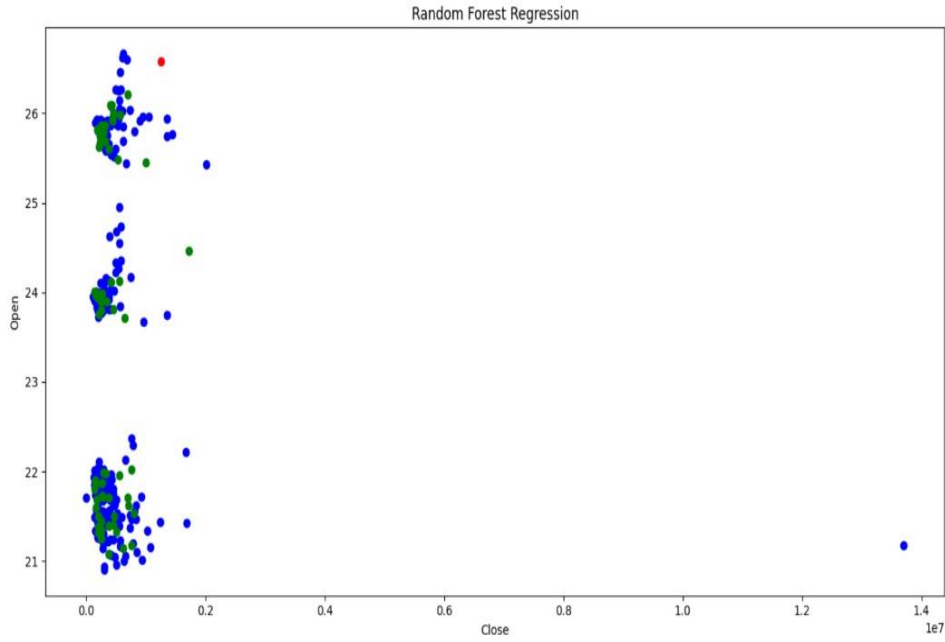


Fig.(i)

To depict the results of stock prices of Uber organization were taken.

A single data point is generated at an interval of 5 minutes where both the algorithm takes five data points in an array at a single time to produce the output.

As shown in the fig.(i) the y-axis shows the prices at which it was opened and x-axis shows prices at which it was closed respective of the time.

Hence for the array of five data point [29.75 29.84 29.72 29.835 29.835] where [26.56920013] was the output. In graphical representation, the figure also depicts the output with a red dot.

For Support Vector Machine Algorithm

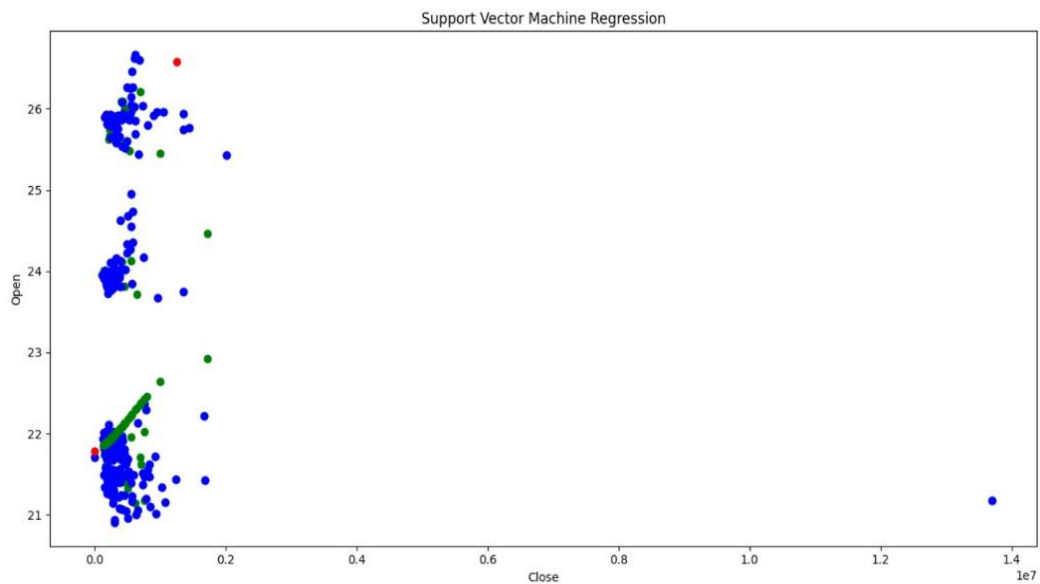


fig.(ii)

Similarly, same input data points of Uber prices were considered and the output predicted was [21.78810464]. Hence the output is highlighted in red dot in graphical representation as shown in fig.(ii)

For average

Therefore, the average of both the algorithm results were calculated to predict the output as [24.18].

6. CONCLUSIONS

A survey on stock market prediction is offered in this study and several techniques and methods for the same have been investigated and analyzed. For better convenience Tkinter GUI was used. For Acquiring accurate predictions of stocks two algorithms were used which is Random Forest Algorithm and Support Vector Machine Algorithm where here the separate results for each algorithm was calculated and the average was taken to get the accurate results. Hence the study stands out because not only single algorithm is used but two different algorithms were used for gaining accuracy which will be more helpful and trustful application in the future.

REFERENCES

- [1] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- [2] Loke.K.S. "Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests", IEEE, 2017.
- [3] Vivek Kanade, Bhausaheb Devikar, Sayali Phadatare, Pranali Munde, Shubhangi Sonone. "Stock Market Prediction: Using Historical Data Analysis", IJARCSSE 2017.
- [4] Sachin Sampat Patil, Prof. Kailash Patidar, Asst. Prof. Megha Jain, "A Survey on Stock Market Prediction Using SVM", IJCTET 2016.
- [5] DivyaYadav1, Dr. N.Thillaiarasu2 (2020) "Stock Market Prediction Using Machine Learning (ML) Algorithms." School of Computing Science & Engineering, Galgotias University, Greater Noida, U.P
- [6] Wasiat Khan Mustansar Ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled H. Alyoubi, Ahmed S. Alfakeeh (2020), "Stock market prediction using machine learning classifiers and social media, news".
- [7] Zexin Hu †, Yiqi Zhao † and Matloob Khushi * "A Survey of Forex and Stock Price Prediction Using Deep Learning"
- [8] Naliniprava Tripathy, "Stock Price Prediction Using Support Vector Machine Approach".