

Stock Market Prediction App Using Big data Analytics

Shashank Nath Yadav¹, Shivam Rai², Shivang Bhatnagar³, Shivansh Srivastava⁴

^{1,2,3,4}Department of Computer Science and Engineering, IMS Engineering College
Adhyatmik Nagar, Ghaziabad, Uttar Pradesh,, INDIA

ABSTRACT

Stock market is very vast and difficult to understand. It is considered too uncertain to be predictable due to huge fluctuation of the market. Stock market prediction task is interesting and complex in nature. Investing in a good stock but at a bad time can have disastrous result, while investing in stock at the right time can bear profits. The aim of this project is to develop a mobile application that will predict the stock market. The approach is to find the best model to analyze historical market data of few select companies and detect hidden relationships between trends in everyday stock prices so as to derive conclusions that may help in predicting the market trends for future.

Keywords: Stock, Predictive model, Android, Firebase, Decision tree, Random forest, Apache pig.

1. INTRODUCTION

Forecasting stock market return is gaining more attention, because of the fact that if the direction of the market is predicted successfully the investors may get benefit a lot. The profitability of investing and trading in the stock market to a large extent depends on several factors. If any system be developed which can consistently predict the trends of the stock market, would make the owner of the system wealthy. More over the predicted trends of the market will help the regulators of the market in making corrective measures [4].

Here we have datasets of different companies listed in New York Stock Exchange (NYSE). A dataset contains a number of features such as Date (D), Opening Price (O), High (H), Low (L), Close Price (C), Volume (V). All the missing values in the dataset are removed from the dataset. Several models are trained on these datasets to predict the stock of companies. Best model here the decision tree is chosen for use in the mobile application.

The stock market is a place which depend on buyers and sellers. When there are more buyers than sellers, the price of stocks increases. When there are more sellers than buyers, the price of stocks decreases. So, there is a factor which causes people to buy and sell the stock. It has to do more with emotion than logic. Because emotion cannot predicted, stock market movements become unpredictable.

The proposed system has not taken factors like emotions, change in company's leadership, internal matters of company, strikes, protests, natural disasters, change in the authority cannot be taken into account when predicting the stock market.

2. CLASSIFICATION MODEL

The following models were used for comparison:

2.1 Decision Tree

In decision tree algorithm we take data set provided as attribute. We give the best data set as input at the root of the node and then we split the tree on the basis of decision that is we split the node on some predefined decision like which splitting would provide us with the best prediction accuracy in our project case we keep on doing the splitting of the nodes until all the leaf nodes of the decision tree are full. We compare all the nodes of the tree with the internal nodes until we reach the leaf node and in the end the tree we receive is our decision tree. The Decision tree that we got for the amazon stock data is below Figure 1.

Decision Tree - Model

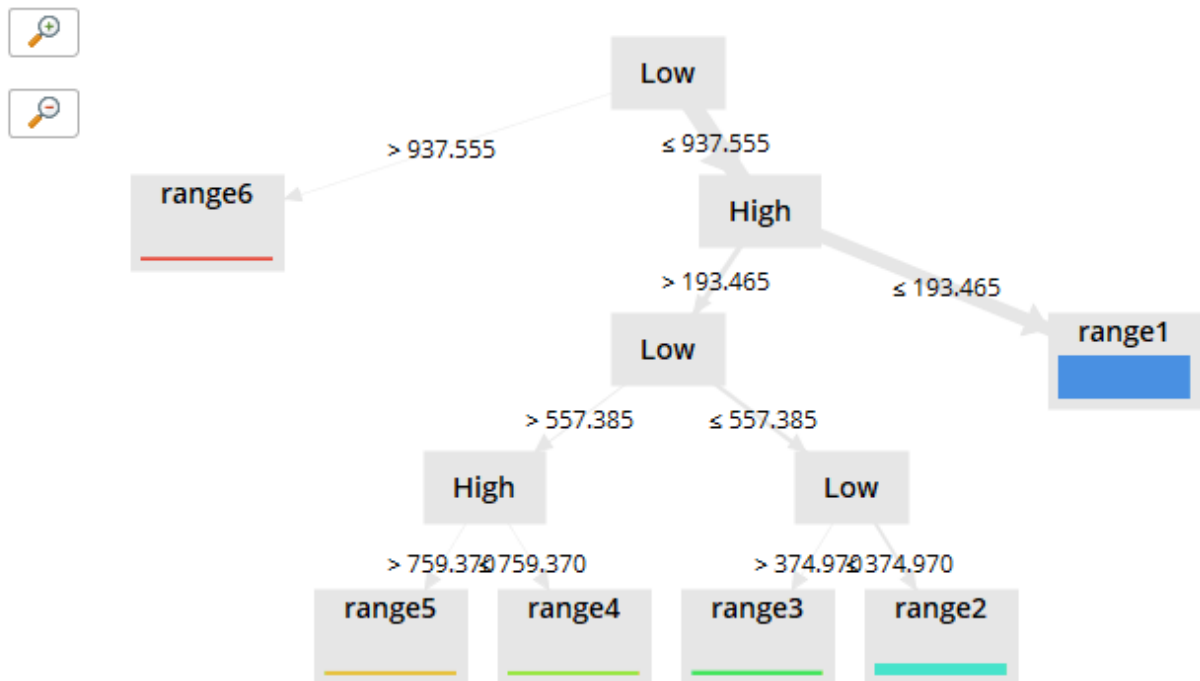


Figure 1. Decision Tree

2.2 Random Forest

Random forests, also known as random decision forests, are a popular ensemble method that can be used to build predictive models for both classification and regression problems. Ensemble methods use multiple learning models to gain better predictive results in the case of Random forest, the model creates an entire forest of random uncorrelated decision trees to arrive at the best possible answer.

3. METHODOLOGY

Three-Tiered Structure:

Back End of the System - Data Mining Platform (SVM) (Developed on Rapid Miner Studio 7.6), Apache pig

Front End of the System – Android application (Developed on Android Studio 3.4)

Integration of the System – Google Firebase

3.1 Apache Pig

We have used apache pig which is high level platform for creating programs to run on Apache Hadoop. The language for the platform is Pig Latin. Pig Latin abstracts the programming from the Java Map Reduce idiom into a notation which makes Map Reduce programming high level, similar to that of SQL for relational database management systems. Pig Latin can be extended using user-defined functions (UDFs) which the user can write in Java, Python and then call directly from the language. By the help of this we refine the data of dataset by removing the missing value and split the dataset into smaller sets as per companies. It also help in extracting the features we want to use to predict the market such as Date, Volume, and Opening Price etc.

3.2 Rapid Miner

We have used this open source free tool for comparing the different algorithms available for prediction purpose using this tool we got the accuracy given by each algorithm for the different stock data that we had provided as input. The auto model was used by us for this purpose and also the standard deviation was kept in mind for selecting the appropriate algorithm. Rapid miner tool also helped us to create a prediction chart.

3.3 Android Application

In our stock market prediction, the android application plays a major role in displaying the future values of the company stock. First the user needs to login for an authentication purpose. The user needs to choose one of the companies for which he needs to know the predicted stock values. The historical data is placed on a private cloud storage. The data is then fed to the rapid miner tool which uses decision tree algorithm to predict the future value of the stock. The data displayed is nearly 99.1% accurate. Therefore this algorithm ensures to provide an accurate data.

3.4 Firebase

Firebase provides a real time database and backend as a service. In our app firebase is used to authenticate the user. This service has also been used to provide a real time database to our application in order to keep track of the already logged in users.

4. RESULT AND ANALYSIS

In this project, different models are used on the datasets to predict the stocks also several factors that are taken into account are Volume Traded, Highest Price, Lowest Price, Opening price and Closing Price.

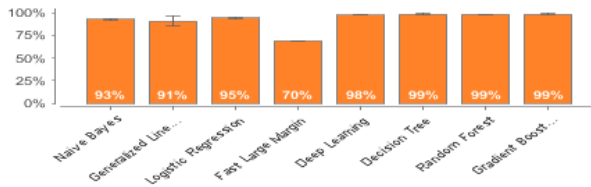
We performed analysis on obtained datasets to establish relation between our output parameters and the selected factors. Fast large margin, deep learning, decision tree, random forest we gave the historical data of a stock as input to the rapid miner tool [5] and then the results [Figure 2] of every model was compared. The result showed that the accuracy given by deep learning was 98%, Fast Large margin was 69.6% Random Forest was 99% and by Decision Tree algorithm was 99.1%.

We came to a conclusion that decision tree model was the best suited model for our prediction purpose with an accuracy of 99.1% though there were other models so we've used decision tree model for our prediction purpose. The prediction made by Decision Tree algorithm is shown in the form of chart [Figure 3]

If we get the stock's previous close price, high, low price for that day we can predict the closing price of the stock for that day. Similarly we could take into account all the news regarding a stock and global market performance to predict the stock's price but that needs further research and study.

Overview

Accuracy



Runtime (ms)



Model	Accuracy	Standard Deviation	Runtime
Fast Large Margin	69.6%	± 0.2%	29 s
Deep Learning	98.0%	± 0.2%	15 s
Decision Tree	99.1%	± 0.5%	6 s
Random Forest	99.0%	± 0.0%	3 min 13 s

Figure 2. Results of different Algorithms applied using Auto model and Compared

Decision Tree - Predictions Chart

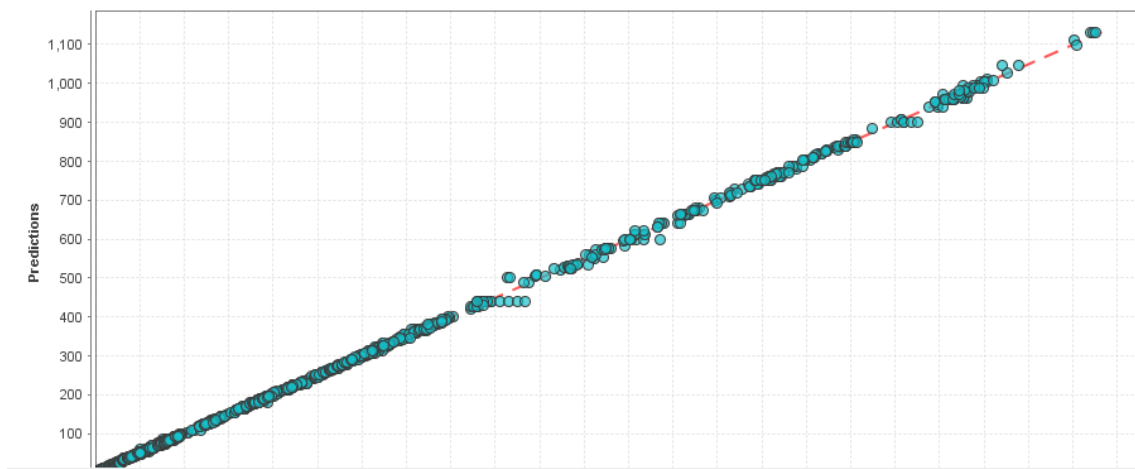


Figure 3. Decision Tree Prediction Chart for Amazon stock data

5. CONCLUSION

Different methods for stock market Prediction were explored and their accuracies was compared. With these results, we infer that the Decision tree Model is more suitable in predicting the stock market. Here it predict the stock with 99.1% accuracy. We also used deep learning and Fast Large Margin models for predicting the closing price of a particular stock but the accuracy provided by them were not up to the mark. So the main confusion was between using Random Forest or Decision Tree algorithm.

Though we got an accuracy of 99% using Random forest method as well but we choose decision tree model as Decision tree algorithm was simpler to implement and also we were more familiar to the decision tree algorithm and also because of a slight better accuracy it had an edge over random forest. Finally, reconsidering the factors affecting the behavior of the stock markets, such as trading volume, news and financial reports which might impact stock price can be another rich field for future studying.

6. REFERENCES

- [1] A Abraham, B Nath, P K Mahanti (2001), Hybrid intelligent systems.
- [2] C D Zopounidis, E M Dimitrakakis, G S Atsalakis (2011). Elliott wave theory and neuro-fuzzy systems, in stock, market prediction: The WASP system *Expert Systems with Applications* 38, pp .9196–9206.
- [3] G S Atsalakis, K P Valavanis (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology, *Expert System Applications* 36, 10696-10707.
- [4] https://en.wikipedia.org/wiki/Stock_market_prediction
- [5] <https://my.rapidminer.com/nexus/account/index.html#home>