

Stock Picker using Machine Learning



Nambirajan M, Rajdev R, Santhosh R, Sharon Raja D, Sivakamasundari G

Abstract: *The main objective of this paper is to build a model to predict the value of stock market prices from the previous year's data. This project starts with collecting the stock price data and pre-processing the data. 12 years dataset is used to train the model by the Random Forest classifier algorithm. Backtesting is the most important part of the quantitative strategy by which the accuracy of the model is obtained. Then the current data is collected from yahoo finance and the data is fed to the model. Then the model will predict the stock that is going to perform well based on its learning from the historical data. This model predicted the stocks with great accuracy and it can be used in the stock market institution for finding the good stock in that index.*

Keywords : *Stock prediction, Machine learning, Stock market, S&P500, Random forest classifier, Long-term trading*

I. INTRODUCTION

THE Stock market is a place, where companies list their share for buying and selling. One can buy or sell stock by registering with the stockbroker. Stockbrokers are intermediators between the investors and the company which owns the shares. The stock market will give huge profits if invested in the right stock. Now a day's stock market has simplified the investing process. So investing in a stock is simple to a person, but predicting the market is very difficult even to financial institutions. Investments in the stock market are of two types. They are intra-day trading and long-term trading.

In intra-day trading, traders hold the stock less than a day whereas in long-term trading traders can hold a stock as long as he wishes to hold the stock. This paper is based on long-term trading. In long-term trading, fundamental analysis is used to find the stock.

Manuscript received on March 15, 2020.

Revised Manuscript received on March 24, 2020.

Manuscript published on March 30, 2020.

* Correspondence Author

Nambirajan M*, Final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering college, Kovilpatti, Tamil Nadu, India.

Rajdev R, Final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering college, Kovilpatti, Tamil Nadu, India.

Santhosh R, Final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering college, Kovilpatti, Tamil Nadu, India.

Sharon Raja D, Final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering college, Kovilpatti, Tamil Nadu, India.

Sivakamasundari G, Assistant Professor(SG), Computer science and engineering, National Engineering College, Kovilpatti, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

There are many parameters associated with the company which specify the company's growth.

Analyzing all the parameters for every company is not possible. This paper deals with automating the stock selection process by using machine learning. 12 years stocks dataset

obtained for yahoo finance is preprocessed first and then it is used to train the model. This model is trained with the help of a Random forest classifier, a machine learning algorithm. Current stock data is downloaded from the yahoo finance and is learned by the model with its experience gained from the training is used to predict the stock growth.

II. EXISTING SYSTEM

In existing system, the market experts select some companies that are well known to all which is listed in the S&P500 index. In such company, they will perform fundamental analysis about the parameters like market cap, enterprise value, revenue, profitability, and its growth, margin and its growth, earning and its growth, matter related to expenses, operating efficiency, dividend payout, cash flow from various activity, short term debt, long term debt, working capital investment, asset growth, etc. Usually, they get this useful information from the annual report of a company, industry-related data, accessing the news and they will calculate the expected growth. Based on the expected growth valued, they will invest in companies.

In this method, one cannot calculate the expected growth of all the companies that are listed in the index. So, some good performing companies may be missed by investors that will give maximum profit in the upcoming year.

III. PROPOSED SYSTEM

The proposed system is developed using machine learning to predict the stock price. This system uses 12 years historical stock dataset download from yahoo finance which is first pre-processed and then it is used for learning. Random forest classifier is a machine learning algorithm, is used for prediction. Then it is back-tested to measure its accuracy. After that, current market data is downloaded and fed to model. The model will predict the stock growth based on the experience gained in the learning process.

IV. METHODOLOGY

A. Data collection

Dataset are collected from yahoo finance and this project contains three datasets they are.

- Historical stock fundamentals
- Historical stock prices
- Historical S&P500 prices.

Historical stock fundamentals

This dataset contains income, future development, return on value, net revenues, and other information to decide an organization's basic worth and potential for future development.

Historical stock prices

The historical stock price dataset contains the price movement of each stock in a day for each company. It mainly contains open, high, low and close data for every single stock.

Historical S&P500 prices

Historical S&P500 price dataset contains the price movement of an index in a day. It mainly contains open, high, low and close data of that index.

B. Data preprocessing

The downloaded S&P500 and stock price datasets do not have data when the market is closed (weekends and public holidays). This creates a gap in the dataset. We need to fill this gap in the dataset for better results.

C. Random forest classifier

Random forest classifier is one of the supervised learning algorithms. It can be used both for regression and classification. This algorithm is easy and flexible to use. A forest consists of trees. It is said that the more tree it has, the sturdier in construction. Random forest classifier creates decision trees based on randomly selected data samples, gets a prediction from each of the trees and selects the best solution by voting. It is a good indicator of the significance of each feature.

D. Random forest classifier working

- Select random samples from a given dataset.
- Construct a decision tree for each of the given samples and get a prediction result from each of the decision trees.
- Perform a vote for each predicted result.
- Select the prediction result with the maximum votes as the final prediction.

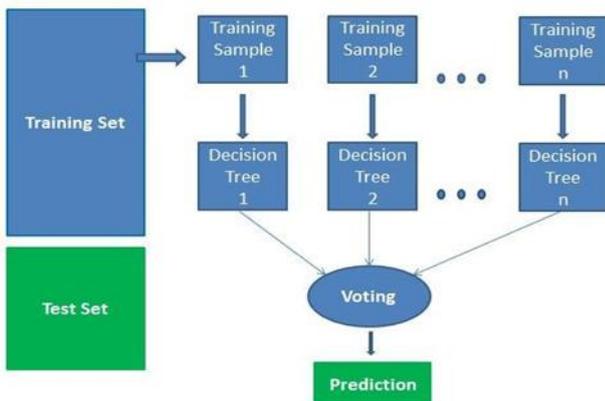


Fig. 1. Random forest classifier.

E. Training

Now the program splits the dataset into a train set and test set, then fits a Random forest classifier to the train set. Then it will print the precision and accuracy of the prediction for the test set.

F. Testing

Testing the model contains two parts.

- Downloading the current dataset.
- Predict the growth of the stock.

G. Current dataset

The current stock data is downloaded and then it is used to predict the stock growth. The current dataset contains current fundamental data of each stock.

H. Predict the stock price

Now the user is prompt to enter the expected growth. Next, the current data are fed to the model to predict the stock price and then it displays the stock that is going to grow more than a specified limit in the upcoming year.

I. GUI based application

A GUI based application is developed with the help of python Tkinter for user convenience. It contains a text box in which users need to enter the expected growth value and then click predict. Now the model will display the output in a Tkinter window.

V. RESULT AND DISCUSSION

A. Screenshots

```

===== RESTART: X:\MachineLearningStocks\stock_prediction
.py =====
Enter percentage of growth you need : 10
Building dataset and predicting stocks...
38 stocks predicted to outperform the S&P500 by more than 10%:
CNX| |IP| |PH| |FDX| |INTU| |WGO| |DGX| |MAC| |RTN| |
LH| |SNA| |GT| |CRM| |BIIB| |GES| |AIZ| |GNW| |VIAB|
|DNR| |LYB| |RL| |ORLY| |R| |TMK| |NWL| |WAT| |SHW| |
PBI| |BLK| |DLX| |HUM| |CME| |GTN| |AMP| |BBBY| |LM|
|EMN| |APD
>>> |
  
```

Fig. 2. Output for 10% growth.

Among the 500 companies in the S&P500 index, 38 companies that are listed above will have a stock value rise by 10% in the upcoming year.

```

Python 3.6.5 (v3.6.5:f59c0932b4, Mar 28 2018, 17:00:18) [MSC v.1900 64 bit (A
D64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: X:\MachineLearningStocks\stock_prediction.py =====
Enter percentage of growth you need : 25
Building dataset and predicting stocks...
8 stocks predicted to outperform the S&P500 by more than 25%:
WGO| |GES| |GNW| |NWL| |PBI| |DLX| |GTN| |BBBY
>>>
  
```

Fig. 3. Output for 25% growth.

Similarly, 9 companies that are listed above will have a stock value rise by 25% in the upcoming year.



Fig. 4. GUI based application.

It is a simple GUI application, it let the user to predict the stock at just one click. At first user need to set expected to growth and then click predict button. Now it will show the prediction. This output is taken for 20% expected growth.

B. Result

Table-I: Expected growth of different companies

S.NO.	Expected growth (percentage)	No. of companies
1	10	38
2	20	9
3	30	6
4	40	1
5	50	2
6	60	0
7	70	1
8	80	0
9	90	0
10	100	0

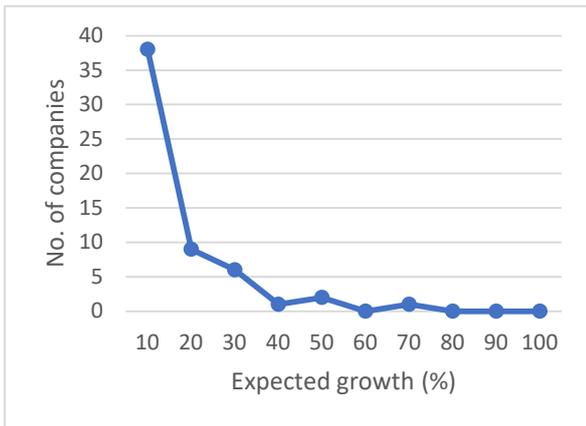


Fig. 5. Expected growth and number of companies.

From the graph it is clear that when the expected growth is increasing then the number companies will be attaining that growth will decrease. So only few companies are good to invest. From the listed companies we can select the best one to invest.

VI. CONCLUSION

A stock market plays a vital role in the growth of the country's economy. This project assists in selecting the stock for investing in the stock market. It can easily be used by the person with the help of GUI based application. This application runs with an accuracy of 79 percent.

REFERENCES

1. Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan, 'Stock Market Prediction Using Machine Learning' Researchgate Conference Paper December 2018
2. M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, 'Stock market prediction using machine learning techniques', 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.
3. K. Raza, 'Prediction of Stock Market performance by using machine learning techniques', 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.
4. Hiransha M, Gopalakrishnan E, Vijay Krishna Menon, Soman K.P, 'NSE Stock Market Prediction Using Deep-Learning Models' International Conference on Computational Intelligence and Data Science (ICCIDS 2018)
5. Sibi P., Jones S. A. and Siddarth P. (2013). 'Analysis of different activation functions using back propagation neural networks'. Journal of Theoretical and Applied Information Technology 47 (3): 1264-1268.

6. Wang J. Z., Wang J. J., Zhang Z. G. and Guo S. P. (2011). 'Forecasting stock indices with back propagation neural network'. Expert Systems with Applications 38 (11): 14346-14355.
7. K. V. Sujatha and S. M. Sundaram, 'Stock index prediction using regression and neural network models under non normal conditions', INTERACT-2010, Chennai, 2010, pp. 59-63.
8. K. A. Althelaya, E. M. El-Alfy and S. Mohammed, 'Evaluation of bidirectional LSTM for short-and longterm stock market prediction', 2018 9th International Conference on Information and Communication Systems(ICICS), Irbid, 2018, pp. 151-156.
9. Python documentation Accessed January 5, 2020 at 11.45 AM. <https://docs.python.org/3/library/tkinter.html>
10. RautSushrut Deepak, ShindelshaUday, Dr. D. Malathi, 'Machine Learning Approach In Stock Market Prediction', IJPAM 2017.

AUTHORS PROFILE



Nambirajan M is a present final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering college, Kovilpatti, Tamil Nadu, India. He did his Higher Secondary from PACM Hr. Sec. School in Rajapalayam, Tamil Nadu, India in the year 2016.



Rajdev R is a present final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering College, Kovilpatti, Tamil Nadu, India, He did his Higher Secondary from Rosemary matric Hr. Sec. School in Palayamkottai, Tamil Nadu, India in the year 2016.



Santhosh Ris is a present final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering College, Kovilpatti, Tamil Nadu, India, He did his Higher Secondary from Karapettainadar Hr. Sec. School in Thoothukudi, Tamil Nadu, India in the year 2016.



Sharon Raja D is a present final year Engineering Graduate in the Department of Computer Science and Engineering at National Engineering College, Kovilpatti, Tamil Nadu, India, He did his Higher Secondary from Rose mary matric Hr. Sec. School in Ramnagar, Tamil Nadu, India in the year 2016.



G.Sivakamsundarii is an Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering at National Engineering College, Tamil Nadu, India. She received her M.E. Computer Science and Engineering degree from PSNA College of Engineering and Technology, Tamil Nadu in the year 2006. She has 18 years of teaching experience. She is an active member in ISTE. She has organized many National Level Seminars and Workshops sponsored by various funding agencies.