

Stock Market Analysis using Machine Learning

¹TS Vinay Kumar, ²Spoorthi Rakesh, ³Yathish C, ⁴Manas Rai, ⁵Ankit Kumar
²Assistant Professor

^{1,2,3,4,5}School of Computing & Information Technology, REVA University, Bengaluru, India
¹vinaykumartsp@gmail.com, ²spoorthirakesh@reva.edu.in, ³C.yathishmbl@gmail.com,
⁴manasrai093@gmail.com, ⁵sinhaankit199@gmail.com

Article Info

Volume 83

Page Number: 4749-4751

Publication Issue:

May-June 2020

Abstract

In this vast finance driven world stock trading is one of the most important and profitable activity. Stock forecast is the strategy for attempting to foresee the future estimation of an organisation stock or other money related resource of an organisation on a trade showcase. People invest in Stock markets based on suggestions which indirectly are based on predictions. Many methods like time-series analysis, statistical analysis and fundamental analysis are used in an attempt to predict but none of them are favourable for perfect prediction due to the volatile nature of the stock market. The programming language used in this context for analysis and prediction is Python and the prediction algorithm used in this context is Linear Regression to predict stock prices for various companies. This paper explains the prediction of stock using Machine Learning.

Article History

Article Received: 19 November 2019

Revised: 27 January 2020

Accepted: 24 February 2020

Publication: 16 May 2020

Keywords: Stock Market, Time series Analysis, Fundamental Analysis, Statistical Analysis, Linear Regression.

1. Introduction

Basically, traders/investors with a lot of money always tend to look for good investment and the best place to invest their money for high profits is the stock market. Quantitative traders always buy stocks at a very small price in high volumes and later on they sell them at a very high price. Stock investigation is a strategy for speculators and merchants to settle on purchasing and selling choices.

There are two essential kinds of stock examination: Fundamental Analysis and Technical Analysis. Fundamental examination focuses on information from sources, including money related records, monetary reports, organization resources and friends' piece of the overall industry. To direct central examination the analysts refer to the company's bank records, balance sheets, cash flow and income statements which are released to the public by EDGAR(database) administered by Securities and Exchange Commission(SEC) and also records released by the company itself. The second kind of stock examination is Technical Analysis. Specialized examination centers around the investigation of over a wide span of time and current stock prices of the company for predicting future values. Fundamental Analysis is more accurate because it considers factors which affect the company's stock value whereas

Technical Analysis does not include the prediction considering the external factors.

In recent years due to machine learning's vast applications and its growing popularity, many traders have started to apply these machine learning techniques for stock prediction due to their efficiency. Stock market is very volatile in nature. The best way to predict tomorrow's value is only by careful analysis. Various technologies have led to the prominent growth of machine learning in predictions. In this project we are going to use one of the most easiest and widely used regression algorithm called as Linear Regression for prediction and various other libraries for effective visualization and analysis.

2. Literature Survey

[1] has used Supervised Machine Learning and Data standardisation for analysis and forecasting. Supervised machine is where the attribute to be predicted is already known. Due to variation of stock data, it is first standardised for easy evaluation and calculation. The Data collected in a CSV format has been normalized and standardised to fit to a whole number with a max difference of 1 unit from its original number. [2]has used Linear Regression strategy for forecasting stock price. In linear regression, facts is modeled by the use of predictor fashions and the difficult to understand attributes are

expected from the prepared information.[3] has used fashionable records mining techniques for wonderful data comparison for desirable structuring of data for convenient understanding. [4] the estimation of the end result parameters. They used the least-square estimation to estimate these parameters. The results have been promising.

[5] offers an assessment of Indian capital market and its structure. In a decade ago or somewhere in the vicinity, it has been resolved that there has been a change in perspective in Indian capital market. The use of numerous changes and propensities in Indian capital market has made the Indian capital market practically identical with the worldwide capital markets.

3. System Design

The Dataset for our project is taken from YAHOO finance website. The dataset is downloaded in the form of a CSV file(Comma Separated Value). Figure 1 represents the data flow diagram for our application . After the data is downloaded and normalized we then proceed to the feature selection where analysis of the dataset is done by plotting graphs and prediction is done using Linear Regression and the accuracy of the predicted data can be analysed by methods of performance measure.

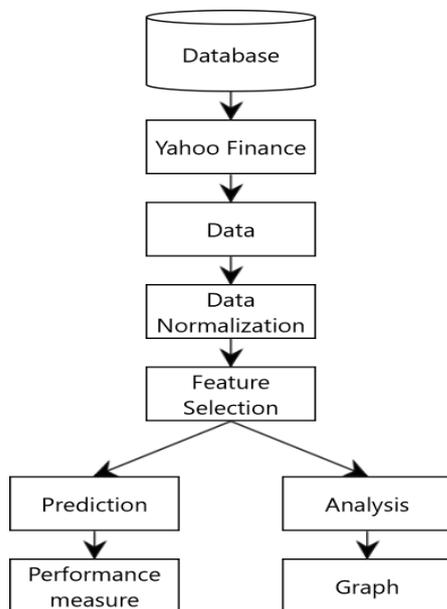


Figure 1: Process Flow

4. Methodology

4.1 Regression:

Regression is a statistical analysis method used in finance, investing and other categories where it aims to establish a relationship between one dependent variable (Unknown value)

Based on a series of other independent variables (known values)

Regression helps money related investors to establish and analyse the relation between various commodities and stocks of business dealing with those commodities. Basically there are two types of regression Simple linear Regression and Multiple linear regression where basic linear regression manages one ward variable(dependent) and one free factor(independent) though multiple linear regression manages one ward variable and a progression of autonomous factors.

4.1.1 Linear Regression

Linear regression is the most easiest and the most widely used regression techniques due to its simplicity. It is a straight methodology for demonstrating the connection between a scalar response (dependent variable) and at least one illustrative variables (independent variable). The connections are demonstrated naturally utilizing the linear predictor model works whose obscure parameters are evaluated from the information. Such models are called Linear Models.

The main idea behind linear regression is to find the best fit for a set of points such that the distance between the points and the best fit line is the least

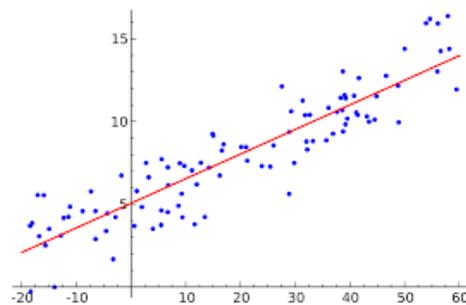


Figure 2: Finding the best fit line

Fig 1 depicts the placement of the best fit curve such that the individual distances between the line and the points is the least. A simple linear regression line has an equation of the form

$$Y = b_0 + b_1X,$$

where X is the logical(explanatory) variable and Y is the needy(dependent) variable. The slant of the line is b, and a is the intercept. Whereas multiple linear regression is represented by the equation

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n,$$

Where Y is the Dependent variable and X₁ , X₂ , X_n are the independent variables.

Linear Regression is extensively used in practical applications .For prediction, forecasting linear regression can be utilized to fit a prescient model to a watched informational index of estimations of both needy and autonomous factors. In the wake of building up the model, if extra estimations of the free qualities are gathered, they can be used to make a prediction of the dependent variable.

4.2 Data Analysis Stage

The dataset is accessed from yahoo finance website automatically over a specified period of dates. The data is stored in CSV file (Comma Separate Values). The attributes of the data include Date, Open (opening price), Close(Closing price), Volume(volume of the stock), Adj Close. The Data is normalized where it is rescaled to the range of 0 and 1.

4.3 Training and Testing Data

The data that has been downloaded and normalized will be used for our machine learning algorithm.

We shall be implementing using Linear Regression method which is the simplest method which is widely used for analysis and forecasting which comes under the Scikit – learn library in python. The data shall be first split into two partitions namely train data and the test data based on certain ratio which in turn depends on the data date range specified.

The training data set shall be used to train the Linear Regression model. The testing data set then shall be used to test the accuracy of the Linear Regression model. Most commonly the ratio of training data set to that of testing data set will be 70% to 30%. As we can see from the figure Below the data is of continuous valued input only as regression works on continuous valued input. The input is given to the Linear Regressor model which is trained and tested. After it is trained and tested our desired value can be input to the trained model which in turn predicts the value for our input.

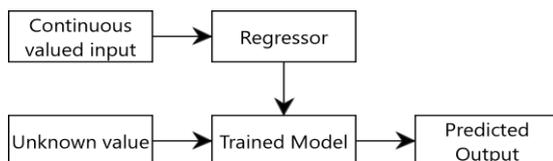


Figure 3: Regression Model

4.4 Results

The accuracy of the model can be calculated by using various Regression metrics which include MEA(Mean Absolute Error) , MSE(Mean Squared Error) and the best fit line can be visualized using matplotlib python library graphs.



Figure 4: Prediction Graph

5. Conclusion

The aim of this project is to help investors and stock brokers for easy and completes up to date analysis and prediction of prices. This application is very easy to set up and operate. Prediction assumes a significant job in business and it's a muddled procedure because of the unpredictable idea of securities exchange.

References

- [1] Kunal Pahwa and Neha Agarwal(2019), " Stock Market Analysis Using Supervised Machine Learning ", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con).
- [2] R.Seethalakshmi(2015), " Analysis of stock market predictor variables using Linear Regression " , International Journal of Pure and Applied Mathematics Volume 119 No. 15 2015.
- [3] Larose, D. T. (2005), "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2,John Wiley &Sons, Inc.
- [4] Ashish sharma ,Dinesh Bhuriya and Upendra singh(2017) "Security analysis and portfolio Management", , I.K. International Publication, 3rd Edition 2008.
- [5] Sameer yadav(2014), "Business forecasting", Prentice Hall of India,Eastern Economy edition,pp.381-457.