International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Implementing Supervised Machine Learning Algorithm to Analyze Stock Market

J. C. Pasalkar[1], Saurabh Kasar[2]

[1]Assistant Professor, Department of Information Technology, AISSMS's Institute of Information Technology, Pune-411001, INDIA
[2]TE. (Information Technology), AISSMS's Institute of Information Technology, Pune-411001, INDIA

A B S T R A C T

The stock market or share market is one of the most important yet complex and sophisticated way of doing business. Small owners, securities companies, banks, everyone we rely on this very institution to generate revenue and share risk. Avery complex model. However, in this paper a machine learning algorithm for predicting future stock prices to use and share open source and existing libraries algorithms that enable this unpredictable business format a little more predictable. Let's see how easy it is the implementation gives acceptable results. Result is it is based entirely on numbers and assumes many axioms may or may not be obeyed in the real world, such as the time of weather.

Keywords: Data Analysis, Fundamental, Implementation, Linear Regression, Stock Market, Supervised Machine Learning

1. Introduction to Stock Market Analysis

For establishment of problem-solving approach through understanding data cleaning, clustering, and classification, data mining is used.The technology is pervasive in nature to summarize and analysis of data from different directions and dimensions. For extracting features from raw data, Machine Learning is very popular. It is also known for being structurally hierarchical and derive prediction. It supports to manage heterogeneous information from the real world with structuring dataset.

1.1 Motivation

- To make use of linear regression to train the classifier which then predicts the value of the label with certain accuracy.
- The accuracy of the predicted value can be checked while training and testing of the classifier.
- Basic knowledge of stock market, graphs and data analysis is coupled with supervised machine learning algorithms for the purpose of stock market analysis.

1.2 Aim and Objective

Aim

This paper aims to analyze stock market and generate accurate predictions using supervised machine learning algorithms.

Objectives

- Predicting the true value of the stock by the method of crowd computing.
- To Examine dependent and independent variables to establish or attempt to establish existing relationships between them.

2. Literature Review

To support this work, several studies have been taken into consideration.

1. STOCK MARKET VOLATILITY – A STUDY OF INDIAN STOCK MARKET:

The analysis showed that amplitude and volatility in bull phase is higher whereas reference period is longer. During expansion, the gains outweigh the loses during the bear phase in a stock market cycle. It also concluded, over the years, declination in volatility can be seen post liberation phase. [2]

2. PREDICTION OF STOCK MARKET DEVIATION USING ARIMA ALGORITHM:

The study discussed about how despite of many resources making stock market investments safer, there still is a myth that it is incomprehensible and complex. They suggested to increase awareness in order for younger generation to better understand pros and cons of investment in stock market. The system implemented consisted of ARIMA algorithm & sentiment analysis in attempt to increase the accuracy of prediction. [7]

3. LINEAR REGRESSION FOR MACHINE LEARNING:

This study explains linear regression on a deeper level. It involves its types, model representation, how predictions are made with linear regression, and how data is prepared for the same.[8]

3. Algorithm

- ***Adj. Close***

This is the information that determines the opening price of the market and for the next day and volume expectancy for the day.

- ***HL_PCT***

This is a derived feature which is defined by

$$HL\_PCT = \frac{\text{Adj. High} - \text{Adj. Low}}{Adj.Close} \times 100$$

- ***PCT-change***

Opens and closes in the same way as High and low because both are very relevant Our predictive model helps reduce i

$$PCT\_Change = \frac{Adj.Close - Adj.Open}{Adj.Open} \times 100$$

- ***Adj. Volume***

this is a very important decision Since the parameter, trading volume is the most direct It influences future stock prices more than other characteristics. So in our case we just use this

4. Experimental Methodology

Data Analysis Stage

- Raw data given is looked upon in-order to identify suitable attributes for the prediction of our selected label.
- The dataset taken is from GOOGL by WIKI and is extracted by quandl using the token "WIKI/GOOGL.
- The attributes of the dataset include:

Open (Opening price of Stock)

High (Highest price possible at an instance of time)

Low (Lowest price possible at an instance of time)

Volume (Total times traded during a day)

Split ratio

Adj. Open

Adj. High

Adj. Low

Adjusted values of above attributes

Adj. Close

Adj. Volume

- After analysing the data and extracting useful information, classifier is used.

Training and Testing Stage

- The extracted data is implemented on the machine learning algorithm.
- SciPy, Scikit-learn and Matplolib libraries in python are used to program our model and train them with the features and label extracted and then test them with the same data.
- After the data is ready to be input in a classifier, the simplest classifier i.e Linear Regression is used, which is defined in Sklearn library of the Scikit-learn package.

Fig. 1. Graph showing stock price of GOOGL from year 2005 till July 2018. Red is the line representing given data and blue is representing the forecasted or the predicted value of stock.

Fig. 2. Graph showing the exact amounts of predicted values.

5. Conclusion

The applied supervised machine learning algorithm, linear regression helped to improvise accurate stock market predictions by data analysis followed by training & testing of raw data provided.

6. References

[1] Andrew McCallum, Kamal Nigam, Jason Rennie, Kristie Seymore “A Machine learning approach to Building domain-specific Search engine”, IJCAI, 1999 - Citeseer

[2] Yadav, Sameer. (2017). STOCK MARKET VOLATILITY - A STUDY OF INDIAN STOCK MARKET. Global Journal for Research Analysis. 6. 629-632.

[3] Montgomery, D.C., Peck, E.A. and Vining, G.G., 2012. Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.

[4] Draper, N.R.; Smith, H. (1998). Applied Regression Analysis (3rd ed.). John Wiley. ISBN 0-471-17082-8.

[5] Robert S. Pindyck and Daniel L. Rubinfeld (1998, 4h ed.).Econometric Models and Economic Forecasts

[6] “Linear Regression”, 1997-1998, Yale University http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm

[7] Agarwal (July 14, 2017). "Introduction to the Stock Market".Intelligent Economist. Retrieved December 18, 2017.

[8] Jason Brownlee, March 2016, “Linear Regression for machine learning”, Machine learning mastery, viewed on December 2018, https://machinelearningmastery.com/linear-regression-for-machinelearning/