

Stock Prediction using Machine Learning

Yash Omer

Department of Information Technology
Bharati Vidyapeeth University, College
of Engineering
Pune, Maharashtra, INDIA
e-mail: yashomer0007@gmail.com

Nitesh Kumar Singh

Department of Information Technology
Bharati Vidyapeeth University, College
of Engineering
Pune, Maharashtra, INDIA
e-mail: nitesh.321.singh@gmail.com

Awadhendra Pratap Singh

Department of Information Technology
Bharati Vidyapeeth University, College
of Engineering
Pune, Maharashtra, INDIA
e-mail: apsingh1096@gmail.com

Dilshad Ashmir

Department of Information Technology
Bharati Vidyapeeth University, College of Engineering
Pune, Maharashtra, INDIA
e-mail: ashmirjack23@gmail.com

Prof. S.Z. Gawali

Department of Information Technology
Bharati Vidyapeeth University, College of Engineering
Pune, Maharashtra, INDIA
e-mail: szgawali@bvuceop.edu.in

Abstract— The paper give detailed on the work that was done using regression techniques as stock market price prediction. The report describes Linear regression methods that were applied with accuracy obtained using this methods, it was found this model is effective from other although there are several opportunities to expand the research further with additional techniques and parameters.

Keywords-linear regression, machine learning, prediction, Support Vector Regression

I. INTRODUCTION

Stock prediction is used to determine the future of company stock or firms. The correct prediction will yield future profit to company's stock. Prediction methodologies [6] encapsulates two philosophies; Fundamental analysis, and Technical analysis. Fundamental analysis, the analyst uses number information such as proportions, earnings to derive future forecast. Technical Analysis, technician utilizes charts and modeling techniques to establish trends in price and volume. Stock prediction is the biggest challenge for experts, researchers and data analysts who are the part of financial sector. This topic has gained many research and analytic groups, achieving the accuracy, confidence and challenges generated. The developed prediction models are applied in the real stock market, if the result were not successful much to beat the real-time market stock, specifically accuracy and generalization point of view. The prediction was performed manually, the experts can reach a measure prediction pattern, but it fails to match the current transaction, could be because of lack of time or confidence, therefore the accuracy is the not our main target, considering prediction time and generalization is our important perspective. High performance Computing (HPC)[6] is also been considered in order to have retrieve "real-time" predictions, the specialized HPC computational resources such as Computational grids are complex and are note easily available for users and financial experts. However, this is an

assumption that we are trying to simplify the model in order to use the chosen regression models.

In this paper, an efficient system for stock prediction is proposed to overcome the challenges, trying to consider prediction time as our crucial measure in addition to the accuracy, using machine learning techniques. In this paper, we have proposed prediction analysis algorithm called Linear regression.

II. PROPOSED SYSTEM

Stock price prediction is a point of issue that has the ability to be worth of dollars and is currently researched by the biggest financial in the world. It is a common problem because it has no clear answer, although many attempts can be made to reach approximation using different machine learning techniques. This project gives opportunity to techniques in real-world machine learning applications analyzing large data sets and using different techniques to train the program to predict the future data.

III. LITERATURE SURVEY

Different methods have been used to predict stock prediction machine learning. Most interesting fields of research include using a type of reinforcement learning called Q-learning [1]. And Import/export growth, earnings from consumers and other data to develop decision tree to determine whether the stock's price will rise or fall. [2]

The Q-learning is machine learning technique which is shown to be effective, but its not clear that how intensive the algorithm will be due to large number of states will be generated. Decision tree is useful when analyzing a specific industry’s growth. The research is also being done on how top-performing stocks are defined and selected and analysis on what can go wrong when modeling stock market with machine learning. [4]

IV. METHODS

A) DATA REPRESENTATION

The data set that used to collect data from Quandl Database [3]is a marketplace with collection of financial, economic data delivered in modern formats used in python Via API. API access is possible through packages for multiple programming such as Python. Data separated in each row consisted of a stock of specific day along with data

on the volume, shares out, closing price and other features for the day.

The python numpy scientific computing library is used along with the data analysis library pandas used to convert CSV files into pandas Data Frame that are indexed by date. Each individual stock is a view of the Parent Data Frame that is filtered based on the stock’s ticker. This gives efficient access to stocks of interest and convenient access to date ranges.

These Data Frame views are then used as the data to be fed into our regression black boxes.

template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation.

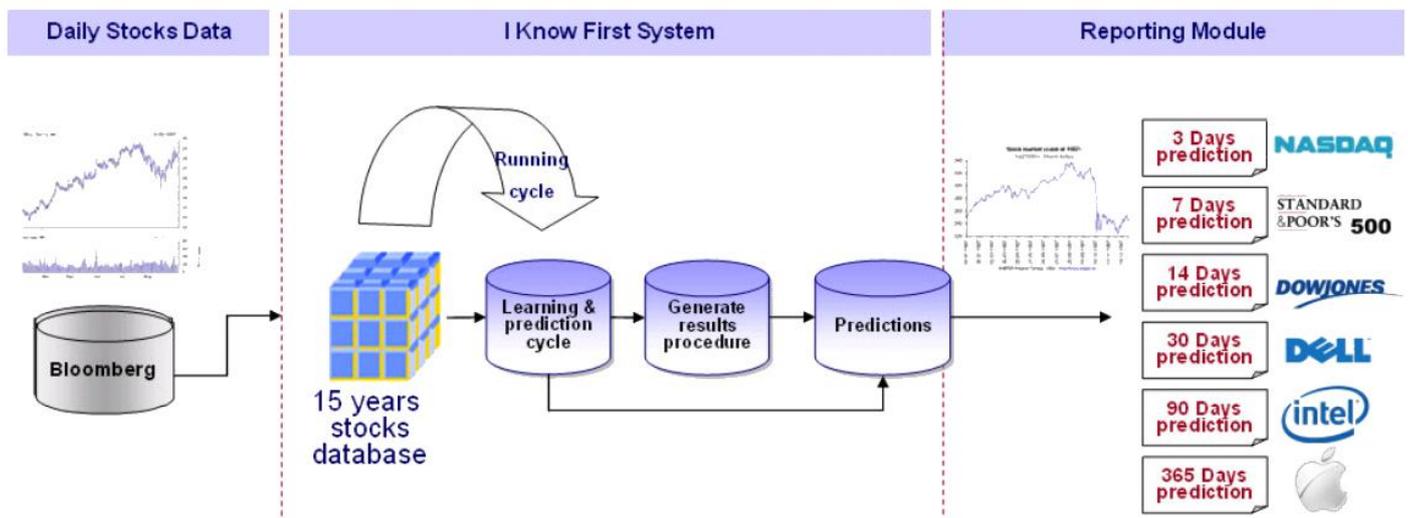


Fig.1: Block diagram of Machine Learning

B) PREDICTION THROUGH REGRESSION

The regression process is done using machine learning scikit-learn library [4]. This is most important part for prediction functionality.

There are some further steps that are used so that data can be fed to into the regression algorithms and return predicted results. The data should have prepared before we fed it in the build model:

a) Data Normalization: Is a rescaling process one or more attributes to range 0 to 1. The largest value for individual attribute is 1 and smallest is 0. Normalization is a technique to use if the data is unknown.

b) Data standardization: It is a process in which data of same type are transformed to one common format that improves comparison process.

C) REGRESSION MODEL EVALUATION

There are many methods for regression that are implemented in scikit-learn [4]. The “sklearn. metrics” module is used which implements several losses, score and utility functions which measures regression performance. Some have been enhanced to handle to multiple output case such as

mean_square_error, mean_absolute_error, explained_variance_score and r2_score.

The multioutput keyword specifies way the scores or losses for each individual target are averaged.

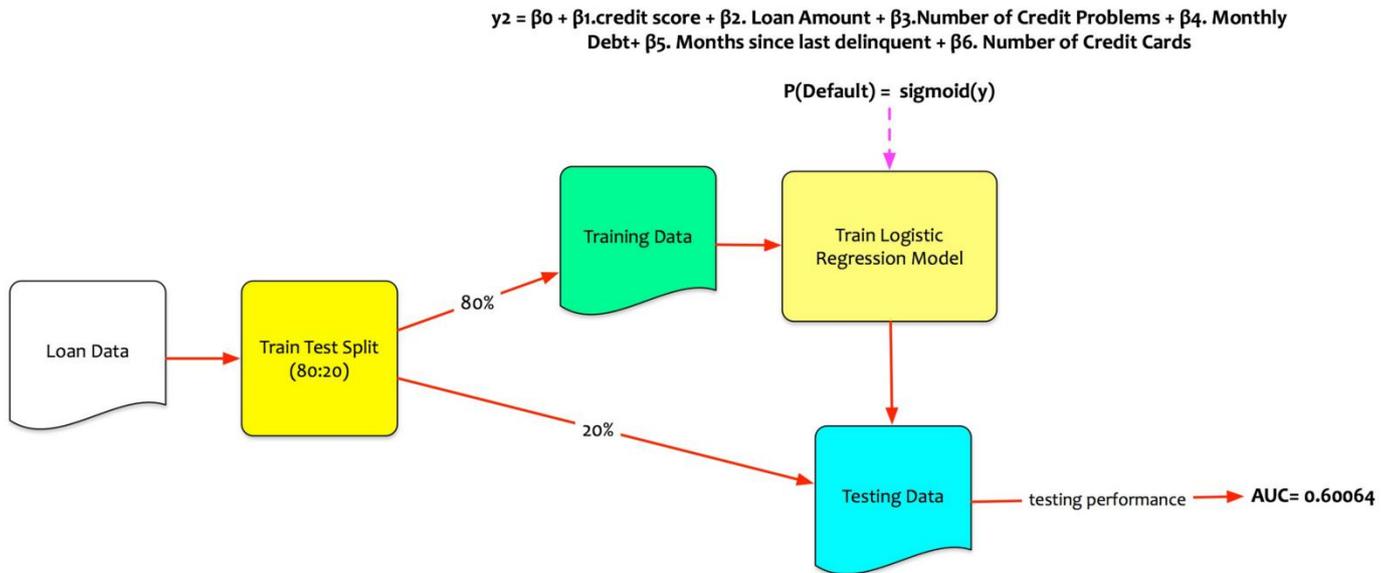


Fig.2: Training Data using Linear Regression

V. RESULTS

Linear Regression [7] is method of making relationship between the dependent variable and independent variable. It is sensitive to normalization technique. The independent variable is called as output variable. The major use for regression analysis are as, it might be used to identify the state of the effect that independent variable had on the dependent variable, it is used to forecast the impact of changes. Regression helps to understand simplest form of

equation containing one dependent and one independent variable is defined by one topic.

where y = dependent variable

regression coefficients

independent variable

$$y = c + b \cdot x,$$

c= constant

b =

x =

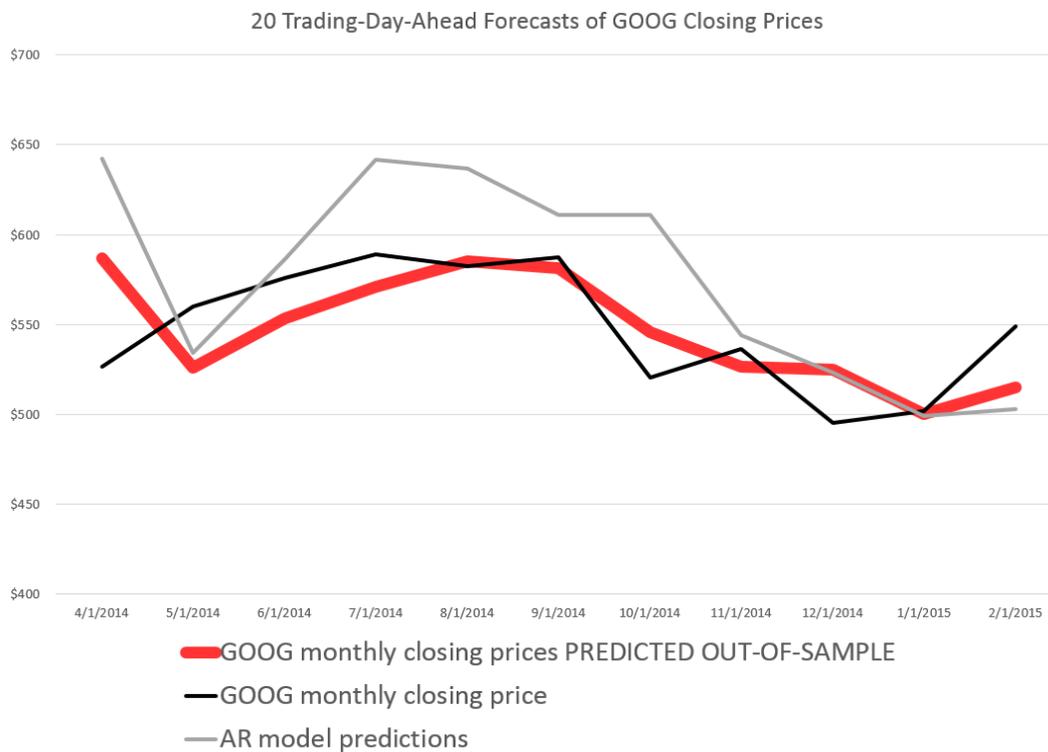


Fig3: Predicting Google Stock

VI. COMPARISON

The comparison version of all the regression is discussed previously. The RBF Kernel in Support Vector Regression shows the better results in overall trails we have run, with

both SVR and Linear Regression., whereas Polynomial kernel varying more as RBF kernel were consistent in short period.

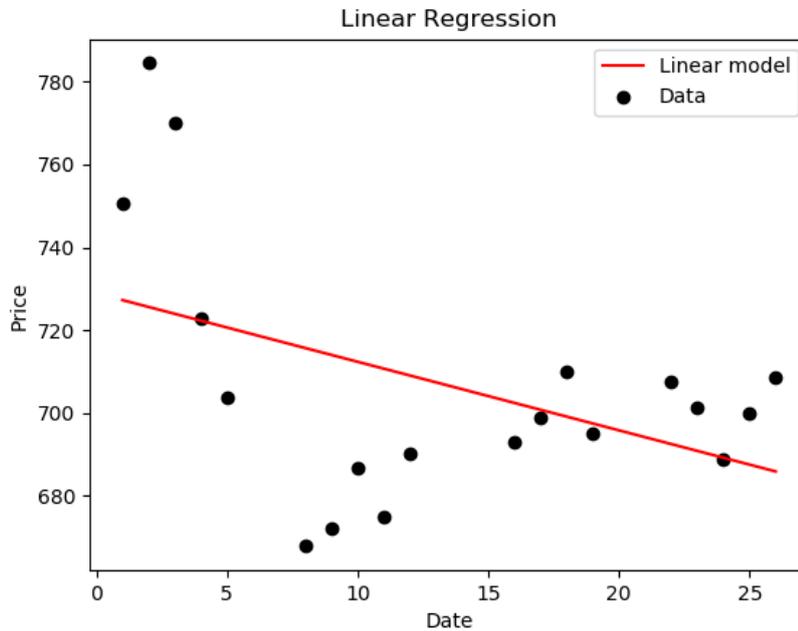


Fig.4: The graph predicts the data of 25 days using Linear regression methods.

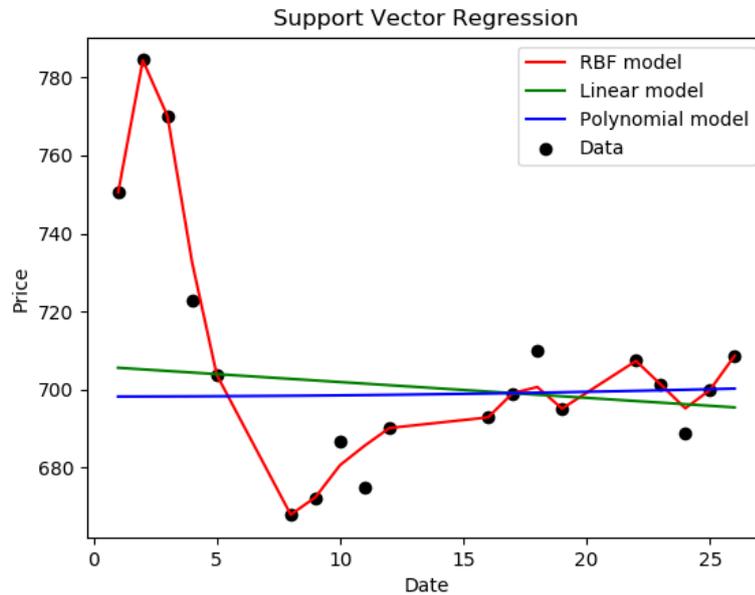


Fig.5: The graph predicts the data of 25 days using Simple Vector Regression methods

It's quite impressive that the most of the trials run in the smaller windows gives better and accurate results, with the prior accuracy.

The result of 28 day's stock price prediction was selected to give the proper insight on the performance of these algorithm for smaller and longer period of time. However, SVR with RBF Kernel performs the best of all, but it was

very impressive to see that SVR with polynomial kernel shows much good results than any other algorithms used.

Linear regression performs shows some poor result when its sizes are small for long-term predictions, then Support Vector Regression. The Linear regression is good for shorter period of prediction using small sizes.

VII. REFERENCES

- [1] R Lee Jae Won, Jonghun Park, Jagmin O, and Jongwoo Lee. 2007, “ A Multiagent Approach to Q-learning for Daily Stock Trading “IEEE TRANSACTIONS ON SYSTEM, MAN, AND CYBERCENTICS 864-877
- [2] C.Tsai and S. Wang 2009, Stock Price Forecasting by Hybrid Machine Learning Techniques. Proceedings of the International Multiconference of Engineers and Computer Scientists, 20-26,, 2009.
- [3] <https://www.quandl.com/data/EOD-End-of-Day-US-Stock-Prices>.
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.
- [5] Hurwitz, E, and T Marwala. 2009. ”Common Mistakes when Applying Computational Intelligence and Machine Learning to tock Market modelling.” University of Johannesburg Press.
- [6] https://en.wikipedia.org/wiki/Stock_market_prediction
- [7] <https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>