# WELCOME

# ANALYSIS OF STOCK MARKET PREDICTOR VARIABLES USING LINEAR REGRESSION

THIS RESEARCH ARTICLE IS WRITTEN BY RAMASWAMY SEETHALAKSHMI

# INTRODUCTION

- History has revealed that stock prices and other resources is an essential part of the important forces of economic activity, and can control or be a pointer of communal mood. In a financial system where the stock market on an increase is measured to be a flourishing economy
- Stock market forecasting is the act of demanding to conclude the future price of a company stock or other financial instrument traded on an exchange. The successful forecast of a stock's future value might give up important profit
- Data mining technique have been effectively revealed to produce high forecasting accurateness of movement of stock price. Now a days, as an alternative of a particular method, traders have to use various predicting methods to increase several signals and more information about the markets future.
- Trade in stock market deals the movement of money of a security or stock from a trader to a buyer.
- Stock market contribution refers to the number of agents who buy and sell equity backed securities either directly or indirectly in a financial trade. Participants are normally subdivided into three distinct sectors; households, institutions, and foreign traders. Direct participation occur when any of the above entities buys or sells securities on its own behalf on a trade. Indirect participation happens when an institutional investor exchanges a stock on behalf of an individual or household. Indirect investment takes in the form of pooled investment accounts, retirement accounts, and other managed financial accounts

# Proposed Algorithm

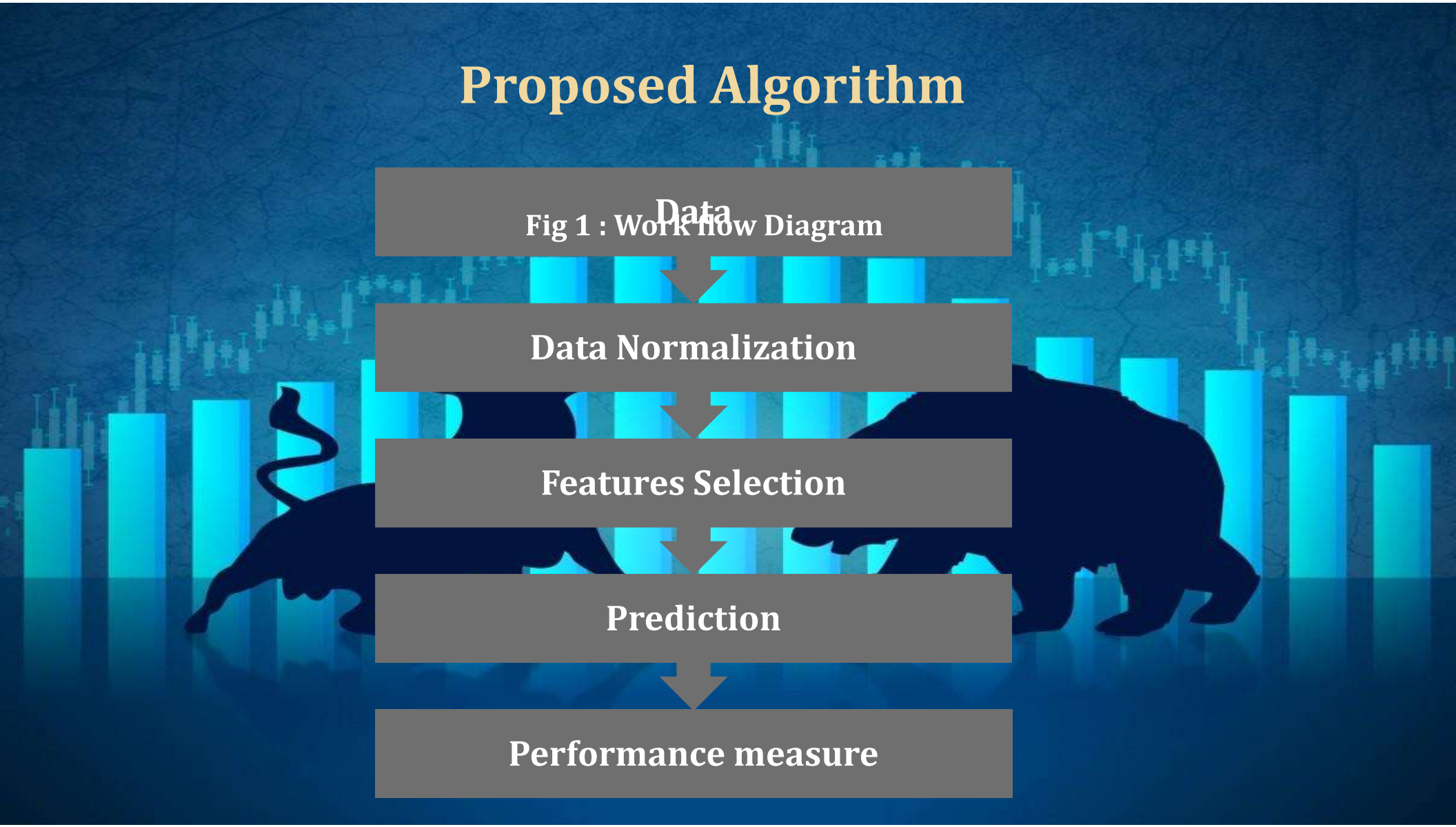


Fig 1 : Work flow Diagram

# Data Description

- The S&P 500index provides the market value of 500 stocks.
- Though it includes stocks from different industries S&P 500 has some distinguishing characteristics such as sensitivity, predictability, scalability ,to name a few.
- We have used all the indices for future calculations.
- A sample of data obtained from [www.yahoofinance.com] is shown in table. Open, High, Low, Close, Volume, Adj Volume are considered as attributes for this study

**Table : 1 Sample Data**

| Date | Open | High | Low | Close | Volume | Adj Close |
|---|---|---|---|---|---|---|
| 6/9/1998 | 1115.72 | 1119.92 | 1111.31 | 1118.41 | 5.64E+08 | 1118.41 |
| 6/8/1998 | 1113.86 | 1119.7 | 1113.31 | 1115.72 | 5.43E+08 | 1115.72 |
| 6/5/1998 | 1095.1 | 1113.88 | 1094.83 | 1113.86 | 5.58E+08 | 1113.86 |
| 6/4/1998 | 1082.73 | 1095.93 | 1078.1 | 1094.83 | 5.77E+08 | 1094.83 |
| 6/3/1998 | 1093.22 | 1097.43 | 1081.09 | 1082.73 | 5.84E+08 | 1082.73 |
| 6/2/1998 | 1090.98 | 1098.71 | 1089.67 | 1093.22 | 5.91E+08 | 1093.22 |
| 6/1/1998 | 1090.82 | 1097.85 | 1084.22 | 1090.98 | 5.38E+08 | 1090.98 |

## Data Normalization

- Firstly, dates are normally represented as strings of the format "YYYY-MM- DD" when it comes to database storage.
- This format must be converted to a single integer in order to be used as a column in the feature matrix. This is done by using the date's ordinal value.

## Features

- Stock market close price is an important piece of information that is very useful for every short-term trader.
- The close prices are very important, especially for swing traders and position traders. It also has implications for practical day trading in many day trading systems.
- The stock market close price level provides very important information about the general mood of investors.
- It tells a lot about the thinking of big investors that allocate large amount of money into the stock market for their asset management purposes.

## Regression Features

- If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of *y* and *X* values.
- After developing such a model, if an additional value of *X* is then given without its accompanying value of *y*, the fitted model can be used to make a prediction of the value of *y*.
- Regression predicts a numerical value .Regression performs operations on a dataset where the target values have been defined already. And the result can be extended by adding new information.
- The relations which regression establishes between predictor and target values can make a pattern.
- This pattern can be used on other datasets which their target values are not known.
- Therefore the data needed for regression are 2 part, first section for defining model and the other for testing model.
- In this section we choose linear regression for our analysis.
- First, we divide the data into two parts of training and testing. Then we use the training section for starting analysis and defining the model.

**Model 1: It includes all the available features . The features are described below.**

- **Opening price**

**The opening price is the value that each share has when the S&P 500 stock exchange opens for trading. The opening price gives a good indication of where the stock will move during the day. Since the Stock exchange can be likened with an auction market i.e. buyers and sellers meet to make deals with the highest bidder, the opening price does not have to be the same as the last day's closing price.**

- **Highest/lowest price of the day**

**The highest and the lowest price of the day are taken the day before and gives an indication of how much the shares usually move during a day and how this in the end will affect the closing price. It also shows the general cyclical movement for each share.**

- **An adjusted closing price**

**It is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next day's open. The adjusted closing price is often used when examining historical returns or performing a detailed analysis on historical returns.**

- **Volume**

**Volume is one of the most basic and beneficial concepts to understand when trading stocks. Volume is defined as, "the number of shares or contracts traded in a security or an entire market during a given period of time."**

## Model 2: In model 2, two features such as volume and adjusted close are omitted

- **Performance measure**

**Models are evaluated through standard performance measures and its description is given in Table2.**

Table 2 Performance measure description

| Name | Formula |
|---|---|
| $R^2$ | |
| Adjusted $R^2$ | |
| F Test | |
| Akaike information criterion AIC | |
| Bayesian Information Criterion BIC | |

# RESULTS AND ANALYSIS

- Six attributes of the Stock data set is considered for Model 1.
- After identifying the most insignificant attribute and eliminate it from data set gives model 2.
- This work compares the model 1 and model 2 using AIC, BIC and $R^2$ values.
- The model outputs are tabulated in Table 3. Model 1 and Model 2 are in Figure 2 and 3 respectively.

Fig 2: Model 1 with all Features

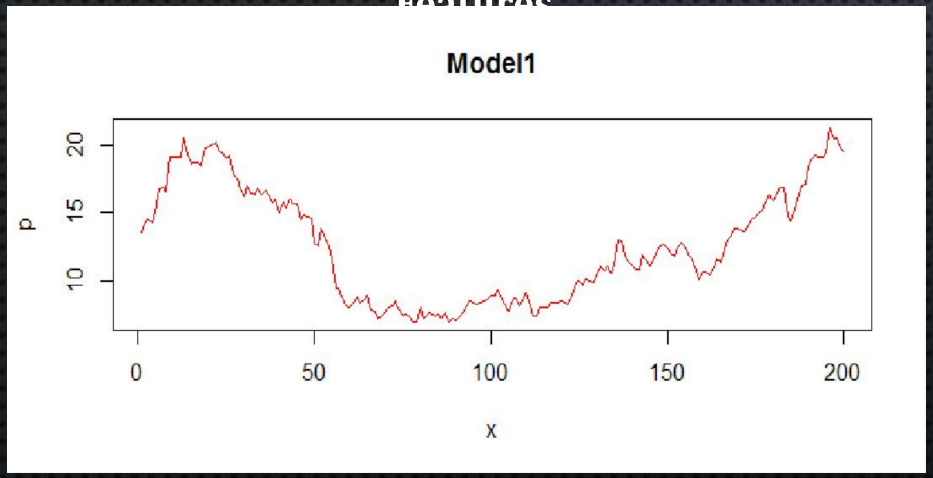


Fig 3 : Model 2 with Selected Features

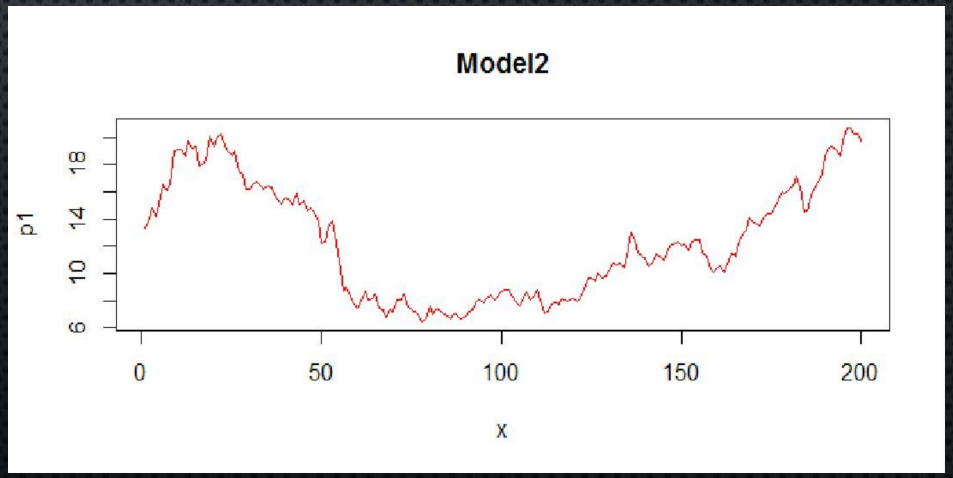

**Table 3 Performance Measures**

| Measures | Model 1 | Model 2 |
|---|---|---|
| Multiple $R^2$ | 0.997 | **0.992** |
| adj$R^2$ | 0.997 | **0.992** |
| AIC | -215.9031 | **3639.1538** |
| BIC | **-183.1108** | **3667.2615** |

- **Model 1 includes all (open, low, high, volume, adj close) attributes and obtained AIC value as: - 215.9031 .Model 2 with the three(open, low, high) attributes and its AIC value is 3639.1538.**
- **Model1 AIC,BIC values are lesser than model2.Hence this work concludes Model 1 is the best model; hence we need to include volume and adjclose attributes for predicting the close price.**

# CONCLUSION

- Model 1 with all features fitted with R Squared value 0.997. This indicates open, high, low, volume and adj close are essential for predicting closing value accurately.
- Model 2 with open, high and low predict close value fitted with R Squared value 0.992 . This indicates prediction of close value is not affected with adj close.
- This study reveals with open, high, low and volume itself enough for finding approximate prediction of close value.

# REFERENCES


- G. E. P. Box and G. M. Jenkins, Time series analysis: forecasting and control. San Fransisco, CA: Holden- Day, 1976. to Bull and Bear Markets, President, Global Financial Data, Inc.
- Draper, N. R., Smith, H., & Pownell, E. (1966). Applied regression analysis (Vol. 3). New York: Wiley
- H.White, Learning in artificial neural networks: a statistical perspective, Neural Computation., Vol. 1 , pp. 425–464,1989.
- Lee, S. 2004. Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. Environmental Management 34(2), 223-232.
- Min, Jae H., and Chulwoo Jeong. 2009. A binary classification method for bankruptcy prediction, Expert Systems with Applications 36(3), 5256-5263.


THANK YOU