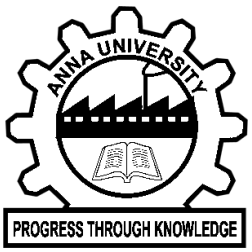
COMPARATIVE ANALYSIS OF VARIOUS REGRESSION MODELS ON STOCK MARKET PRICE PREDICTION

CA5314 – MACHINE LEARNING TECHNIQUES LABORATORY
MINI PROJECT REPORT

by

LAKSHMAN RAJ G (2021179023)



Department of Information Science and Technology
College of Engineering Guindy Campus (CEGC)
Anna University Chennai

JANUARY 2023

ABSTRACT

Stock market plays a very important role in fast economic growth of the developing country like India. Performance of Stock market and country growth is tightly bounded with each other. The number of people engaging themselves with the stock market investment is very less because of the riskiness on the capital attributed due to the dynamic nature of the stock market. Though people invest in stock market based on some prediction. To predict the stock market prices people search methods and tools which will increase their profits, while minimizing their risks.

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price will maximize investor's gains. Stock market prediction is a major challenge owing to non-stationary, blaring, and chaotic data, and thus, the prediction becomes challenging among the investors to invest the money for making profits. Several techniques and models have been proposed and devised to predict the stock price prediction in the past.

There are various types of regressions which are used in data science and machine learning. Each type has its own importance on different scenarios, but at the core, all the regression methods analyse the effect of the independent variable on dependent variables. This project aims to perform a comparative analysis on various regression models in predicting the price of a stock.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The financial market is a dynamic and composite system where people can buy and sell stocks over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Inspired by the growing popularity of machine learning algorithms for predicting applications, these algorithms could be used to uncover hidden patterns in the trajectory of stock prices. Most markets now have historical information readily available. One can use these historical values to study and analyse how a particular stock has been performing. This data can be studied using various Machine Learning Algorithms to predict the stock price movements.

1.1.1 Problem Description

Stock market prices are so volatile, and it depends on various factors. Machine learning algorithms can be employed to predict the stock prices to some accuracy. In machine learning there are various regression models to predict the independent variable movement based on the dependent variable. In this paper the accuracy of these models is compared in predicting the stock prices using the historical data.

1.2 OBJECTIVE OF THE STUDY

- To apply various regression models on the historical data of stock prices.
- To predict the future price of stock using regression models.
- To compare the accuracy of each model in predicting the stock prices.

CHAPTER 2

METHODOLOGY

2.1 GENERAL

Figure 2.1 deals with the methodology followed in the project.

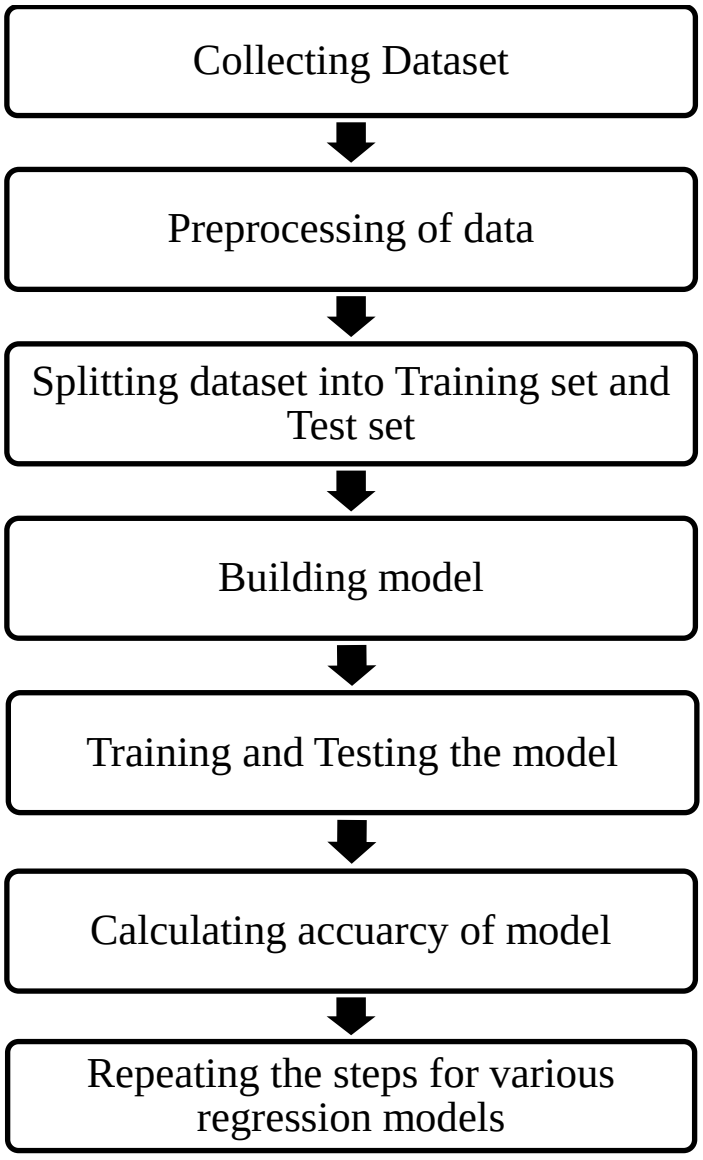


Fig 2.1 Methodology followed in the project

2.2 COLLECTING DATASET

The dataset for the project is collected from the National Stock Exchange (NSE) website and from Yahoo Finance. The NSE website gives the list of Nifty 50 stocks. The yfinance module in python offers a way to download historical market data from Yahoo finance.

2.3 PREPROCESSING OF DATA

The dataset contains features like Open, High, Low, Close, Adjusted Close and Volume of individual stocks. The open, high, low, close prices are the main data points from the technical analysis perspective.

Open – When markets open, the first price at which a trade executes.

High – The highest price at which the share transacts for the given day.

Low – The lowest price at which the share transacts for the given day.

Close – The Close price is the most important price because it is the final price at which the market closed for a particular period.

Adjusted Close - The closing price after dividend pay-outs, stock splits, or the issue of additional shares.

Volume - The total number of shares traded in a specified time frame. This would include every share that is bought and sold during the period.

To avoid the problem of overfitting and underfitting of model, last 100 days of data is chosen for model construction. The Open, High, Low price of the stock act as the independent variable(X) and the Closing price of the stock is the dependent variable(y).

2.4 SPLITTING DATASET INTO TRAINING SET AND TEST SET

The dataset is split into Training and test set using train_test_split function from sklearn python module. The 75% of the dataset is taken for training the model while the rest of the data is kept for testing the model.

2.5 BUILDING MODEL

The models used in this project are the regression models which are available in sklearn module. The regression models that are employed are Multi-Linear Regression, Decision tree Regression, Lasso Regression, Bayesian Regression and Polynomial Regression.

2.6 TRAINING AND TESTING THE MODEL

The various regression models are trained using the training dataset. The models are then tested with the test dataset to predict the Stock Closing price.

2.7 CALCULATING ACCUARCY OF MODEL

The parameters Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) are used to define the accuracy of the model.

RMSE is a metric which ranges from 0 to infinity, where the closer the score is to 0 the better performing the model is. R-squared values range from 0 to 1, the higher the R-squared, the better the model fits your data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}} \qquad R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

CHAPTER 3

RESULTS AND DISCUSSIONS

3.1 TEST RESULTS

The pyplot function in matplotlib module is used to visually analyze the model results. The Stock chosen for the study is Reliance Industries Ltd.

Fig 3.1 – 3.5 gives the results of each model prediction.

The multi linear regression model has RMSE of 15.52 and R^2 of 0.89.

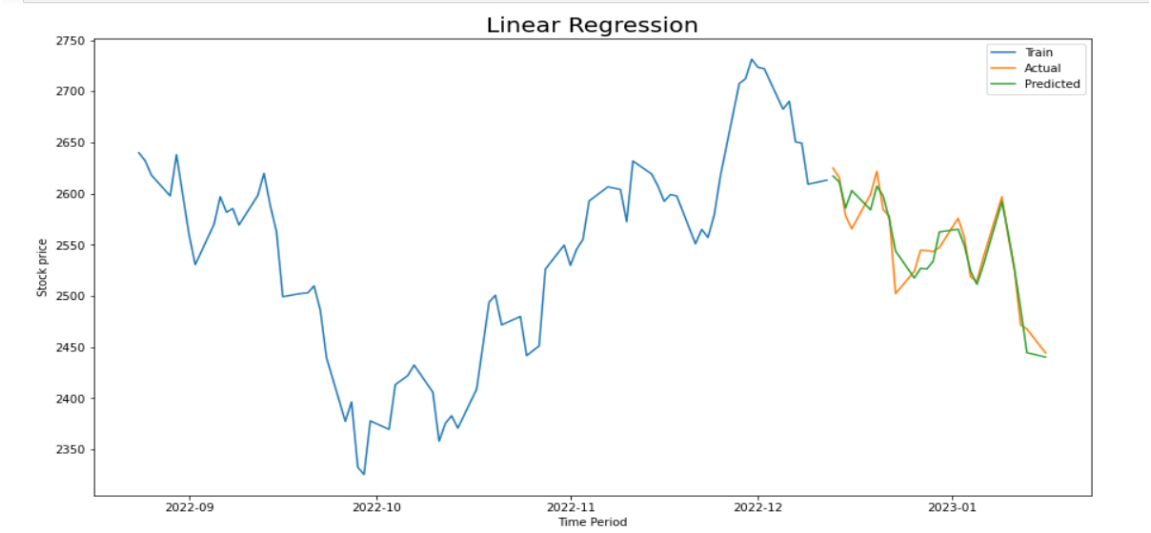


Fig 3.1: Plot of Multi Linear Regression Model

The Decision Tree regression model has RMSE of 24.13 and R^2 of 0.73.

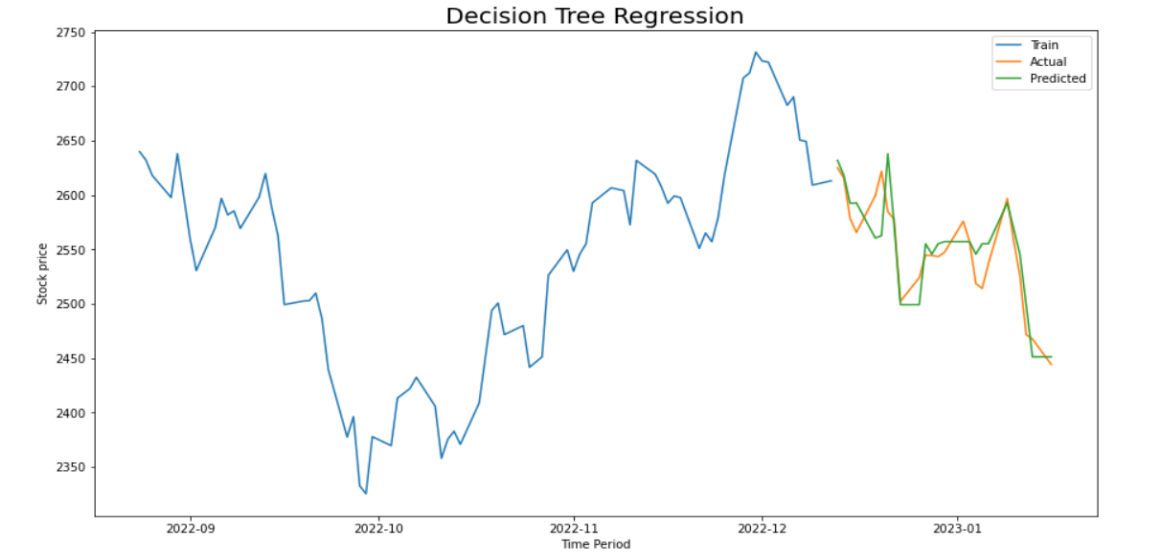


Fig 3.2: Plot of Decision Tree Regression Model

The Lasso regression model has RMSE of 15.55 and R^2 of 0.89.

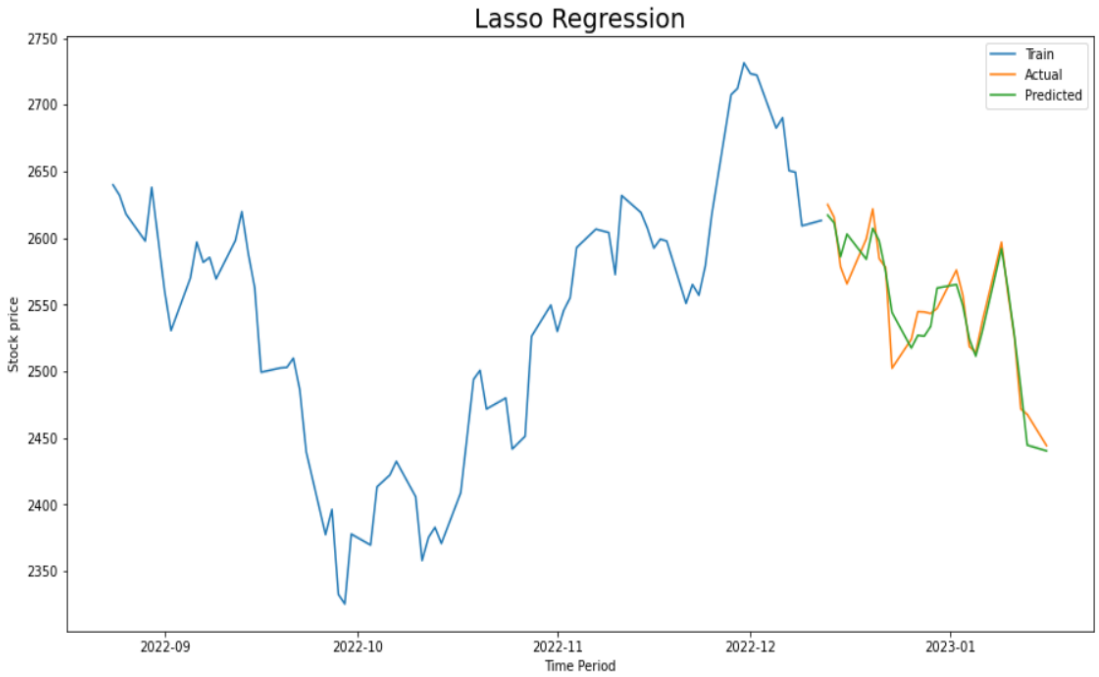


Fig 3.3: Plot of Lasso Regression Model

The Bayesian regression model has RMSE of 15.54 and R^2 of 0.89.

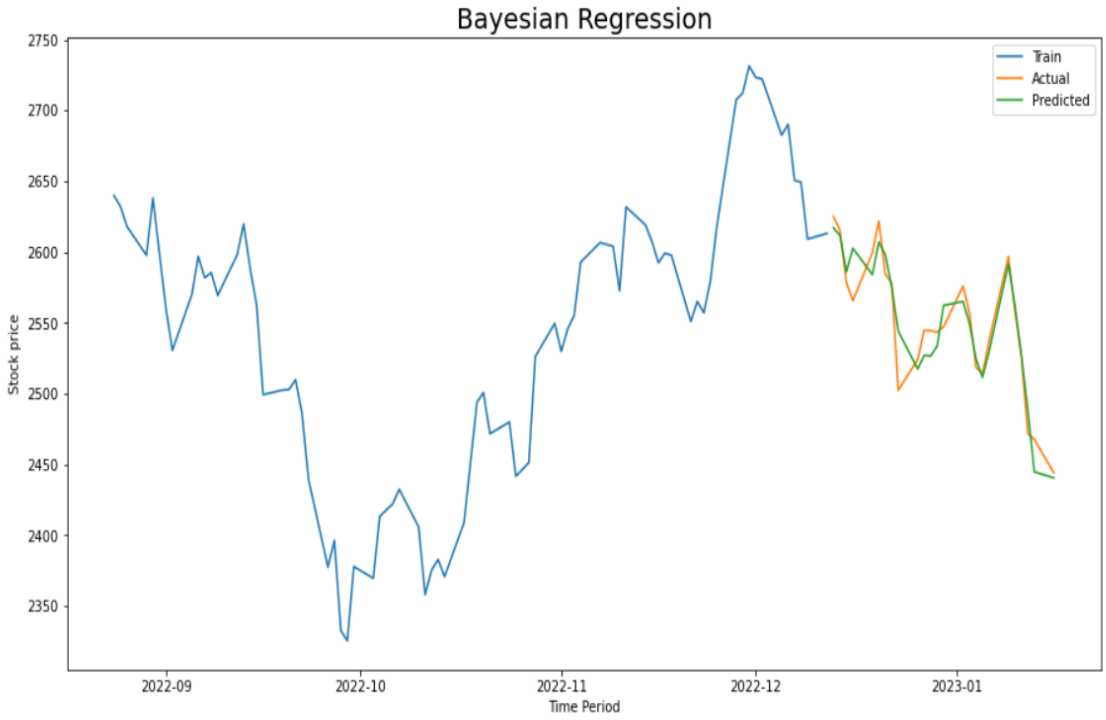


Fig 3.4: Plot of Bayesian Regression Model

The Polynomial regression model has RMSE of 13.18 and R^2 of 0.92.

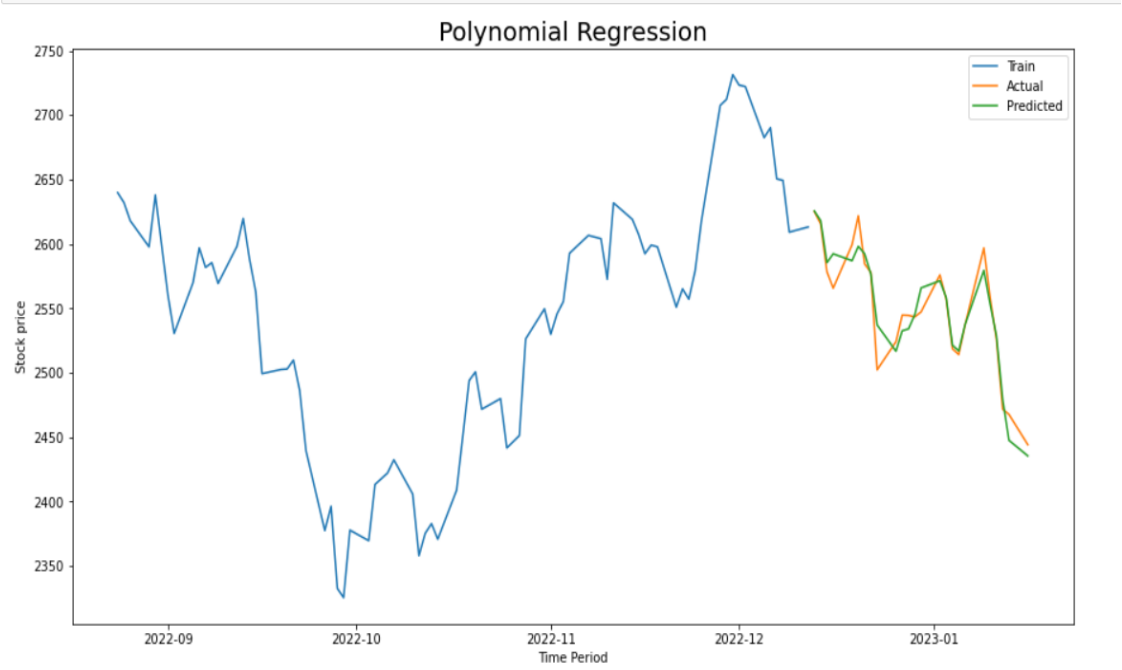


Fig 3.5: Plot of Polynomial Regression Model

The following table 3.1 gives the accuracy metrics of each model.

Table 3.1: Accuracy metrics of models

MODEL	RMSE	R^2
LINEAR REGRESSION	15.52	0.89
DECISION TREE REGRESSION	24.13	0.73
LASSO REGRESSION	15.55	0.89
BAYESIAN REGRESSION	15.54	0.89
POLYNOMIAL REGRESSION	13.18	0.92

CHAPTER 4

CONCLUSIONS

The results obtained from various regression models have been tabulated and the following conclusions are drawn.

- Regression algorithms are applied on the pre-processed datasets.
- Using the regression models, the price of stocks are predicted and compared with actual stock prices and it is visually represented.
- From the plotted graph, the variation in predicted and actual stock prices can be interpreted easily for each of the regression models.
- On comparing the accuracy metrics of models, it has been found that Polynomial Regression model with degree of 2 gives better accuracy with lesser RMSE value of 13.18 and higher R^2 value of 0.92.
- On the other hand, Decision Tree Regression model has lesser accuracy with higher RMSE value of 24.13 and lesser R^2 value of 0.73.
- The above-mentioned conclusions are arrived only for the historical data of Reliance Industries Ltd shares.

CHAPTER 5

REFERENCES

1. Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020), “A survey on machine learning for stock price prediction: algorithms and techniques”, University of Southampton, Southampton, UK.

2. Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A., (2020), “Stock closing price prediction using machine learning techniques”, Jaypee Institute of Information Technology, Noida, India.

3. Parmar, Ishita, et al., (2018) "Stock market prediction using machine learning.", first international conference on secure cyber computing and communication (ICSCCC), IEEE, 2018.