Stock Closing Price Prediction System

Name:	**B Yuvaraj Goud**
Registration No./Roll No.:	19071
Institute/University Name:	IISER Bhopal
Program/Stream:	DSE
Problem Release date:	February 02, 2022
Date of Submission:	April 24, 2022

1   Introduction

A stock market is a place where investors can buy and sell ownership of investible assets. It is the most profitable investment sector in the field of the financial market. Predicting stock prices is the most challenging task in financial market analysis. Making precise inferences about the future price of the stock market has apparent benefits. In this project we are trying to implement various machine learning techniques on data collected from 01/01/2014 to 01/01/2016 of 88 stocks from Yahoo Finance to predict the closing price. The given data set consists of 97732 rows × 5 columns. Below figure is plot between opening price and data targets of given data set.
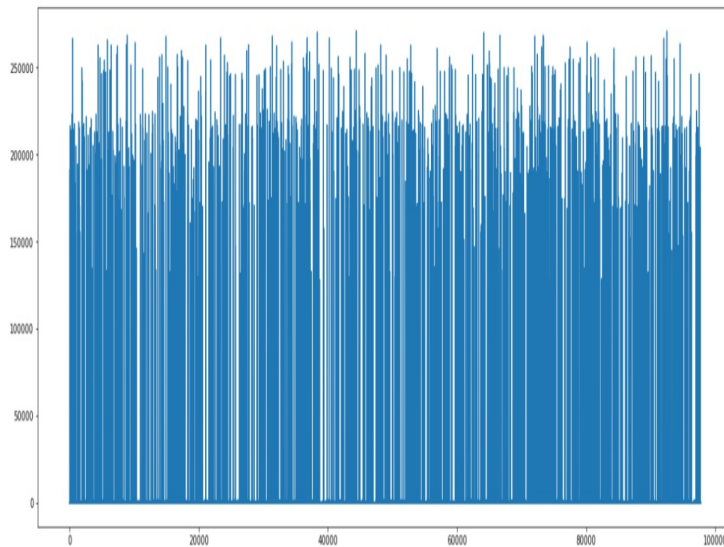

Figure 1: Open vs data targets

2   Methods

2.1   Getting our data

- Imports: Importing necessary libraries such as pandas, numpy, seaborn, matplotlib, sklearn etc.,

- Fixing our data: After the data was loaded into our IDE i.e. Jupyter notebook, three missing values were identified in each column and they were dealt by excluding them from the data frame as it showed best performance when compared to replacing them with mean/median/mode.

2.2 Linear Regression

- Importing LinearRegression, metrics, mean-squared-error, r2-score and seaborn from sklearn.

- Fitting our data: Splitting the data into training and testing data and trained the model using testing data.

- Fitting the training data into the linear regression model.

- Printing the prediction results based on the trained model.

2.3 Support Vector Regression

- Importing SVR, StandardScaler, MinMaxScaler, GridSearchCV from sklearn.svm, sklearn.preprocessing, sklearn.modelselection respectively.

- Transformation of features by scaling each feature of the data.

- Fitting our data: Splitting the data into training and testing data and trained the model using testing data.

- The performance of the model is then improved by using parameter tuning, setting kernel = 'rbf', C = 1e3, gamma = 0.1

- The closing prices are then predicted and Mean squared error is calculated.

2.4 Decision Tree Regression

- Importing of necessary libraries from sklearn.tree such as DecisionTreeRegressor

- Splitting the data into training and testing data and trained the model using testing data and tree score is calculated on test data.

- Closing prices are then predicted on the tested data.

- Model Coefficients, Mean Absolute Error, Coefficient of Determination are then calculated.
  Github link : https://github.com/Yuvaraj7UV/STOCK-CLOSING-PRICE-PREDICTION-SYSTEM/blob/

3 Evaluation Criteria

3.1 For Linear Regression:

R2 score : 0.99998, MSE: 7.5608, Model coefficient: [[-5.92196679e-01 8.19573559e-01 7.74535152e-01 1.51590073e-06]]

3.2 For Support vector Regression:

Model Coefficients: [[-0.59078429 0.81888992 0.77216585]], Mean Squared Error: 0.00977921124615093, R2 score: -0.5468888713528417

3.3 For Decision Tree Regressor:

Model Coefficients: [[-0.59078429 0.81888992 0.77216585]], Mean Absolute Error: 3.9821150467533e-05, Coefficient of Determination: 0.9999653284773713.

4 Analysis of Results

Amongst the used models Linear Regression model gives best results of closing price of the given data set. Thus this model can be used to predict the closing prices which can benefit the investors to maximize their profit and minimise the loss.

5 Discussions and Conclusion

Linear regression is the analysis of two separate variables to define a single relationship and is a useful measure for technical and quantitative analysis in financial markets. Plotting stock prices along a normal distribution—bell curve—can allow traders to see when a stock is overbought or oversold.

The Disadvantages of Linear Regression Linear Regression Only Looks at the Mean of the Dependent Variable. Linear regression looks at a relationship between the mean of the dependent variable and the independent variables. Linear Regression Is Sensitive to Outliers. Data Must Be Independent.

Stock market prediction means determining the future scope of market. A system is essential to be built which will work with maximum accuracy and it should consider all important factors that could influence the result. Various researches have already been done to predict stock market prices.

References

1. https://www.kaggle.com/yash612/stockmarket-sentiment-dataset
2. https://github.com/kartik-joshi/Stock-predection/blob/master/AAPL-2.csv
3. https://www.bseindia.com/indices/IndexArchiveData.html