Mobile Price Prediction

Name:	**Abhishek Prasad**
Registration No./Roll No.:	20011
Institute/University Name:	IISER Bhopal
Program/Stream:	EECS
Problem Release date:	February 02, 2022
Date of Submission:	April 16,2023

1 Introduction

Prediction of Closing Price of stocks is of great interest for day traders, swing traders and active investors around the globe. Correctly predicting closing price of stock can make good sum of wealth from the market. For the same here I am using data of two years of 88 companies of different sectors to predict the closing price of stock any particular day. Data consist of four features:

"Open": Opening price of the Stock

"High": Highest price reached by the Stock in a day

"Close": Price of Stock when Stock Exchange got closed.

"Volume": Number of Shares Traded in that particular day.

All the features are numerical. No categorical feature.

2 Methods

2.1 Data preprocessing

In this step, we performed data cleaning by eliminating all the null values present in the dataset. We also reshaped the data to its correct form so that it can be used for further analysis.

2.2 Models Used For Regression

This problem is a regression problem as the target variable is continuous. For regression we are using:
- Linear Regression
- Random Forest Regression
- Support Vector Rgression

2.3 Hyperparameter Tuning

Table 1: Hyper-parameters of different models

Models	Hyper-parameters Space
Random forest regression	• n estimators: [100,200,300] • max depth: [7, 9,11] • min samples split': [6, 8, 10] • min samples leaf: [2, 3, 4]
SVR	• C : [1.0,1.5,2.0] • gamma: [0.1, 1, 'scale', 'auto'] • kernel:('linear','rbf','polynomial','sigmoid') • epsilon: [0.01, 0.02, 0.025]

We used grid search cv to tune the hyper-parameter of the chosen models. After running grid search cv we found the best results for random forest regression with n estimators 200, max depth 8 and for svr with linear kernel, gamma 0.3. For linear regression there is no hyperparameter tuning. After getting the best parameters for each model we ran a k-fold cross-validation with 10 split to get the mean square error. I also used a statistical method in which i predicted the closing price by taking the average of high and low value of stock price.

3 Feature selection

We did the feature selection by finding the p value of each feature. By doing so we found the volume feature to have p-value more than 0.05 so we removed that feature and kept other feature for further analysis.

4 Analysis of Results

Table 2 shows the mean square error and r2 for all the regression models used in this project

Table 2: Performance Of Different regression models

Regressor	MSE	R2 value
Random forest regression	8567	0.99999744
Linear Regression	5527	0.99999867
SVM	48307.18	0.9995496
Statistical method	7789.18	0.9999596

5 Discussions and Conclusion

We found that linear regression gave the best result for predicting the closing prices of the stocks.Link to the code GitHub link