Yes Bank Stock Closing Price Prediction

Rajat Chaudhary, Anukriti Shakyawar,

Raman Kumar, Deepmala Srivastava

Team: Web Crawlers

Abstract: Yes Bank is an Indian bank headquartered in Mumbai, India and was founded by Rana Kapoor and Ashok Kapoor in 2004. It offers wide range of differentiated products for corporate and retail customers through retail banking and asset management services. On 5 March 2020, in an attempt to avoid the collapse of the bank, which had an excessive amount of bad loans, the Reserve Bank of India (RBI) took control of it. RBI later reconstructed the board and named Prashant Kumar, former chief financial officer and deputy managing director of State Bank of India, as MD & CEO of Yes Bank, along with Sunil Mehta, former non-executive chairman of Punjab National Bank, as Yes Bank's non-executive chairman.

Introduction

Yes Bank has interests in Retail, MSME and Corporate banking. It has three subsidiaries – YES Securities (India) Limited, YES Trustee Limited and YES Asset Management (India) Limited. As of September 2018, Yes Bank had taken syndicated loans from eight large international entities including ADB, OPIC, European investment bank, banks in Taiwan and Japan for amounts ranging from US$30 million to US$410 million. It also partnered with the US government based OPIC and with Wells Fargo to support women entrepreneurs.

Yes Bank provides (UPI) Unified Payments Interface facility to allow customers to easily and securely perform various financial transactions from their mobile devices via third-party app providers like PhonePe and Yuva Pay. According to the data shared by NPCI (National Payments Corporation of India), Yes Bank processed 25.94 million transactions amounting to INR 14811.73 crores through its own UPI app in July 2021.Yes Bank acquired over 24.19% stake in Dish TV, India's largest direct-to-home (DTH) company in terms of subscribers, on 30 May 2020.[32]

1.Data Description

DATASET:

In this Dataset We have 285 Rows and 5 Columns namely Date, Open, High, Low and Close.

1) Open : Opening price of the stock of particular day

2) High : It's the highest price at which a stock traded during a period

3) Low : It's the lowest price at which stock traded during a period

4) Close : Closing price of a stock at the end of a Trading Day

5) Date : We will use it as an index

1.Analysis Methodology

Integral research, data cleaning and filtering, data visualization, data transformation, that make up for our three-part analysis strategy. We started by performing some fundamental research on our dataset. When we did this, we found the basic information regarding our data set such as columns, data types, shape, info and we also want to found in dataset if there is some missing values, duplicate values present or not but Fortunately We found no duplicate values, no missing values. Thirdly We have done data visualization. In Data visualization we plot graph between independent variables vs date, dependent variables vs date, relation between dependent variables and independent variables.

3.1. Data Cleaning

During data cleaning we found no missing values, no null values and no duplicate values. In pandas info operation We found Date as object datatype and hence we convert object datatype to date datatype. Also we have seen much variation between within feautres.

3.2. Data Visualization

In our Data Visualization we performed many analysis to find relations in our data set. First we plot closing price vs Date.

Fig.1: Trend of Yes Bank Closing Price

From the above graph we found closing price falls after 2018 which is a sign of worried for investors.

Fig.2: Distribution of Closing Price

From this plotting we found that Closing price is right skewed. It may lead us to misleading results in view of statistical hyposthesis. It can be corrected by applying Log Transformation then we'll have a look how this data behave.

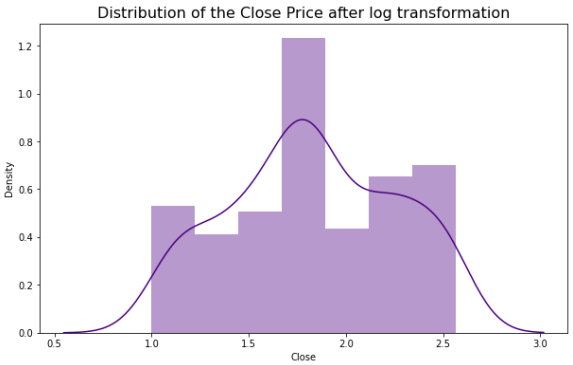
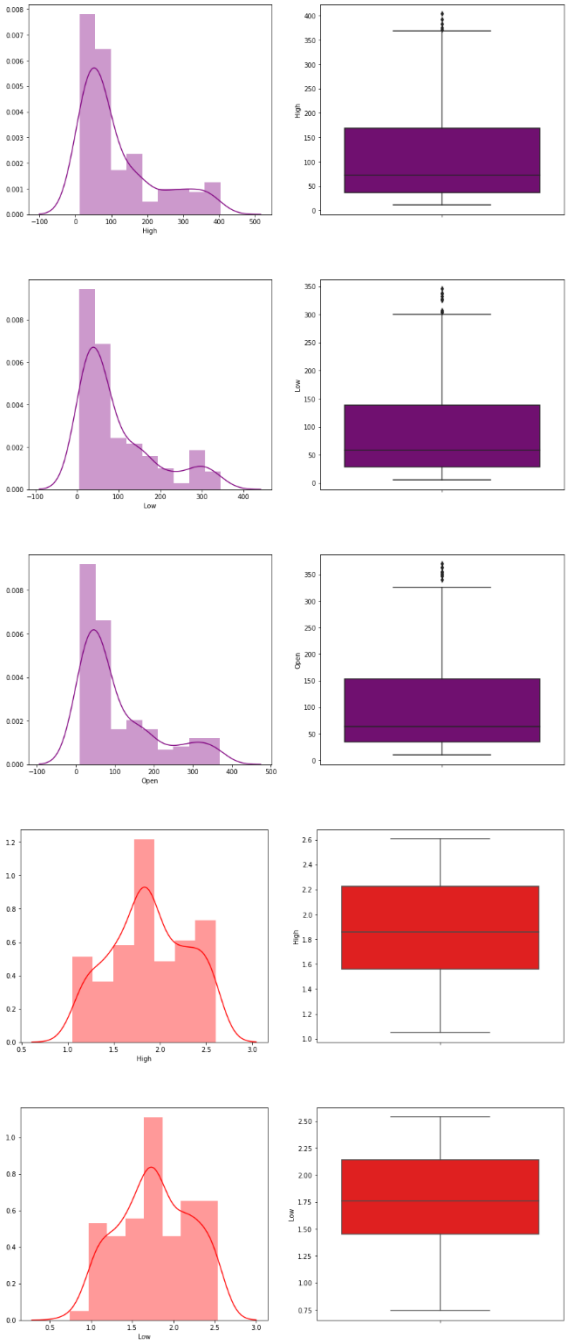


Fig.3: Distribution of the Close Price after log transformation

This garph shows the distribution of closing price is uniform not right skewed or left skewed. This graph is achieved by log transformation of previous graphs.



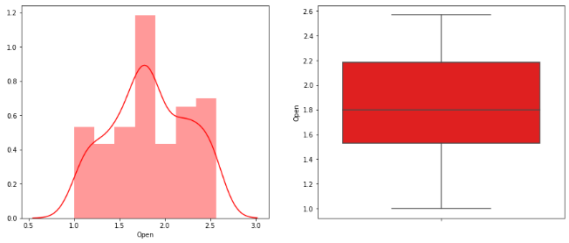
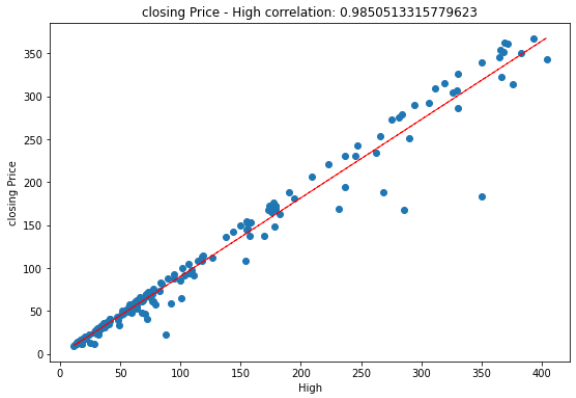
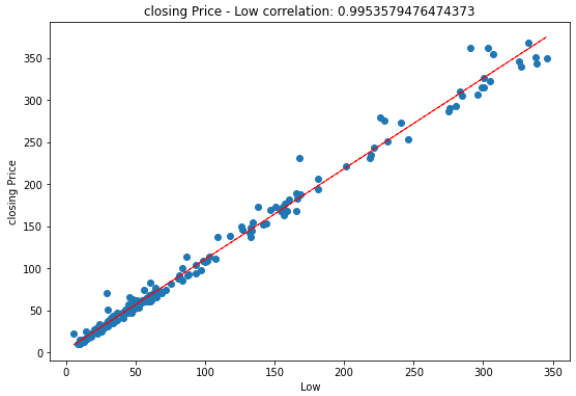


Fig.4: Distribution of High, Low & Open before and after log transformation

The same procedure of data tranformation we applied in all features like High, Low and open in order to achieve uniform distribution respectively.





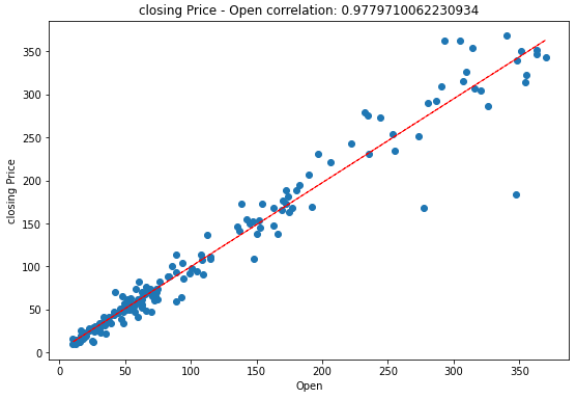


Fig.5: Correlation between Independent variables and Dependent variables

From the above graph we can see that there is linear relation and high correlation between each independent variables and dependent variables.

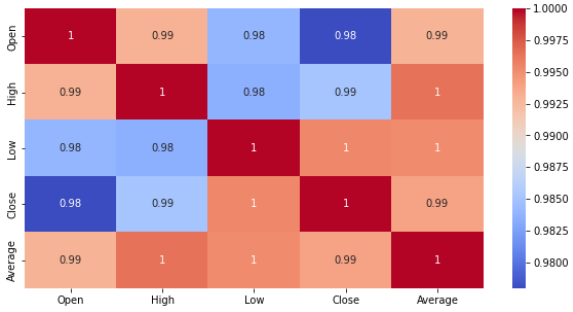


Fig.6: Heat Map for correaltion between variables

There are very high correlation between independent variables which lead us to multicollinearity. High multicollinearity is not good for fitting model and prediction because a slight change in any independent variable will give very unpredictable results. To check multicollinearity and how much it is in our dataset, we have to calculate VIF(Variation Inflation Factor) so, we can decide which varible we should keep in our analysis and predicting model and which should be removed from the datasets.

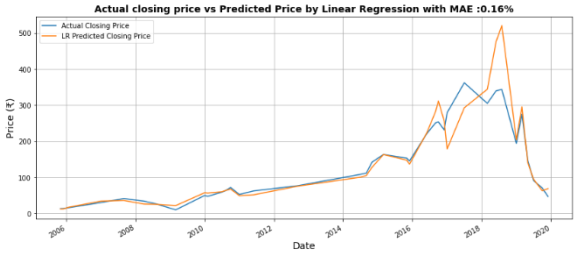


Fig.7: Actual closing price vs Predicted Price by Linear Regression

This graph predicts closing price with training accuracy of 94.03% after introducing dummy variables and removing columns which is highly correlated and that leads to multicollinearity.

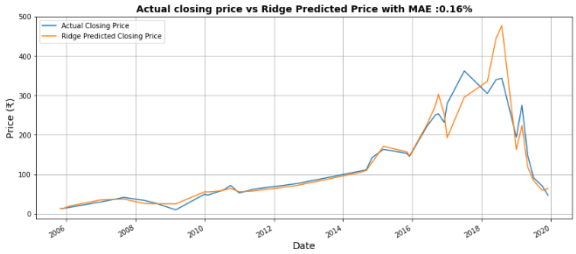


Fig 8. Actual closing price vs Predicted Price by Ridge Regression

This graph predicts closing price with training accuracy of 94.58% after hyperparamter tuning and cross validation and also after introducing dummy variables and removing columns which is highly correlated and that leads to multicollinearity.

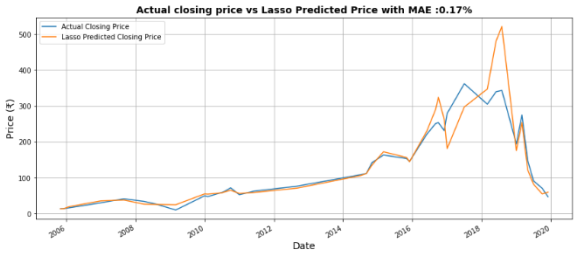


Fig 9. Actual closing price vs predicted price by Lasso Regression

This graph predicts closing price with training accuracy of 94.58% after introducing dummy variables and removing columns which is highly correlated and that leads to multicollinearity.

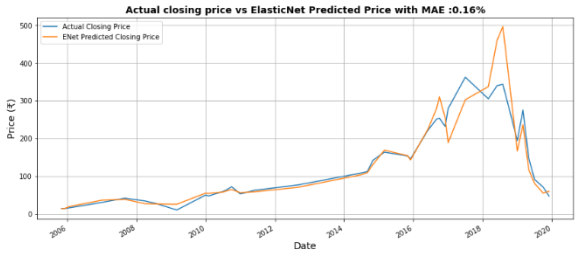


Fig 10. Actual closing price vs predicted price by ElasticNet Regression

This graph predicts closing price with training accuracy of 94.58% after introducing dummy variables and removing columns which is highly correlated and that leads to multicollinearity.

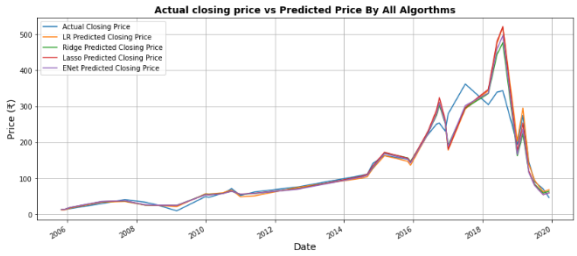


Fig 11. Actual closing price vs predicted by all algorithms

In this graph we can see that most of the apps with higher rating are smaller apps in size, we also put paid and free apps in this also but the type of apps are evenly distributed so the type of the app does not affect the rating but the size of the app surely does.

Conclusion

1. Target Variable is strongly dependent on Independent Variables.
2. Linear Regression and Lasso are performing better than other models with training accuracy 94.0359% and 94.45777% respectively.
3. Apart from Linear Regression and Lasso, Ridge and Elastic Net is also performing better but they have less training accuracy.
4. Ridge and ElasticNet is performing far much better after Applying Hyperparameter Tuning and Cross validation, it is because we have small set of datasets.
5. R2 and Adjusted R2 are around 95 and 91% in each model.

Future Work

We can explore hyperparameter tuning and cross validation in order to gain more accuracy and also find ways to reduce multicollinearity.

References

1. https://github.com/ramank123/Yes-Bank-Stock-Closing-Price-Prediction/blob/main/data_YesBank_StockPrices%20(1).csv

2. https://pandas.pydata.org/

3. https://matplotlib.org/stable/api/index.html

4. https://scikit-learn.org/stable/

5. https://seaborn.pydata.org/examples/index.html