Yes Bank Stock Closing Price Prediction

MANISH GUPTA,
Data science,
AlmaBetter

Abstract:

Yes Bank is a well-known bank in the Indian financial domain. Since 2018, it has been in the news because of the fraud case involving Rana Kapoor. Owing to this fact, it was interesting to see how that impacted the stock prices of the company and whether any predictive models can do justice to such situations.

The main objective is to predict the stock closing price of the month.

Keywords: *Dependent and Independent variables, Distribution.*

1. Problem Statement

Yes Bank is a well-known bank in the Indian financial domain. Since 2018, it has been in the news because of the fraud case involving Rana Kapoor. Owing to this fact, it was interesting to see how that impacted the stock prices of the company and whether Time series models or any other predictive models can do justice to such situations. This dataset has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month

2. Introduction

The monthly stock prices of YES BANK since its inception and includes closing, starting, highest, and lowest stock prices of every month is provided. The main objective is to predict the stock closing price of the month.. Then I have to look for the Null values and outliers in the data, separating the dependent and independent variables , then I have to check the skewness of the data, Then I have to import linear regression model and metrics that we will use for evaluating different models performance then I have to use LASO, Ridge and Elastic-Net regression with cross validation technique. After that I have to generate few related charts and then testing the result

3. Exploring the database

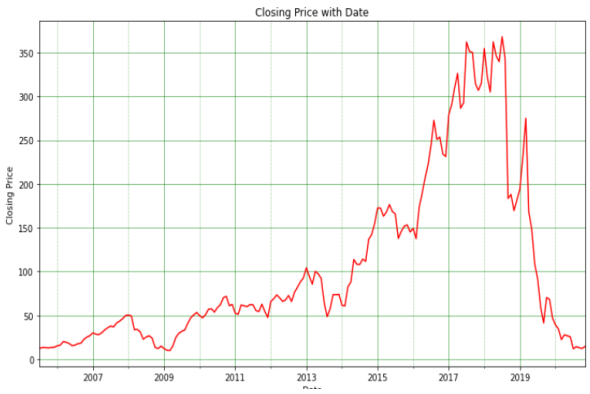
We have provided the database

- Shape of this database is (185,5).
- There are five columns.
- The graph demonstrates how closing price varies with each passing year
- We can clearly see from the graph that around 2018, when the fraud case involving Rana kapoor came to light, a clear significant dip can be seen in the stock price of Yes Bank data

Explaining the data:- We have a dataset containing values of Yes bank monthly stock prices as mentioned in our problem statement.

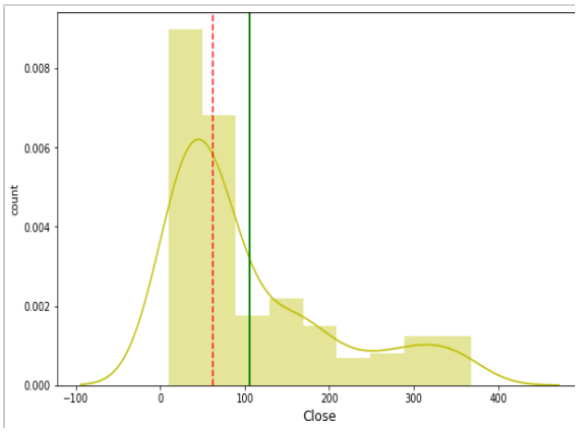
Explaining the features present :-

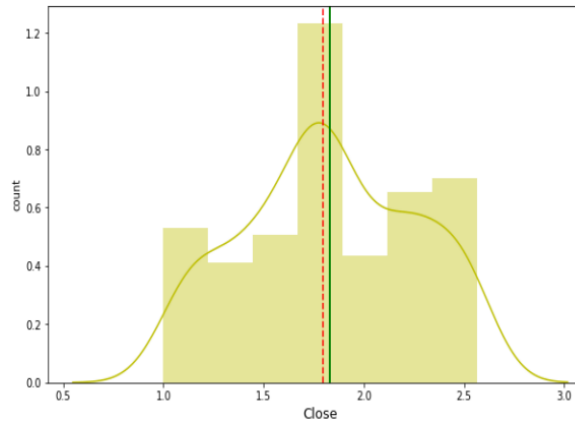
- Date :- The date (Month and Year provided)
- Open :- The price of the stock at the beginning of a particular time period.
- High :-The Peak(Maximum) price at which a stock traded during the period.
- Low :-The Lowest price at which a stock traded during the period.
- Close :- The trading price at the end (in this case end of the month).



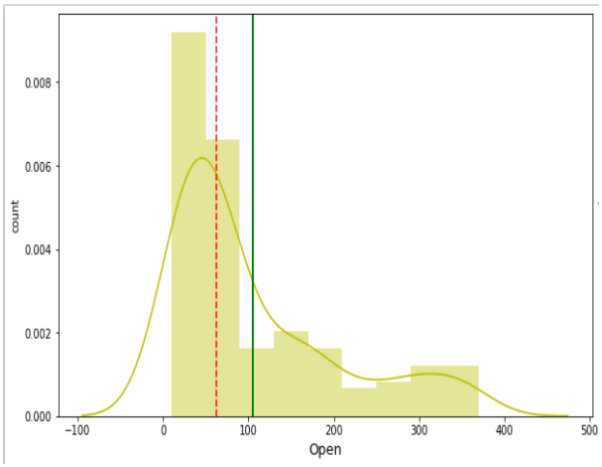
4. Variable analysis

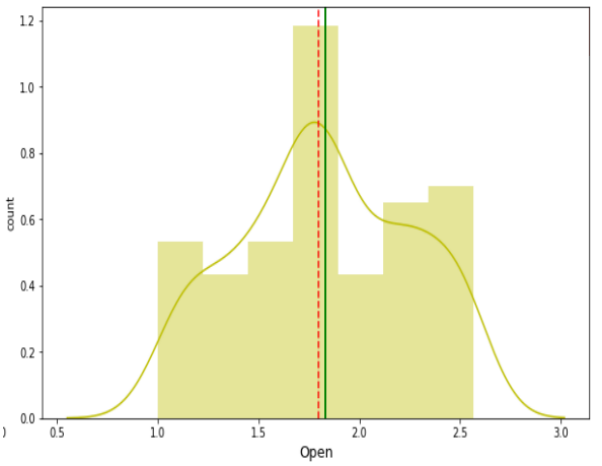
Plotting the dependent variable. We can see that our dependent variable close is positively skewed . So we do a log transform on it and plot it as seen in the right chart. This makes it approximate normal distribution and is optimal for our model's performance. Now our mean and median are nearly equal.

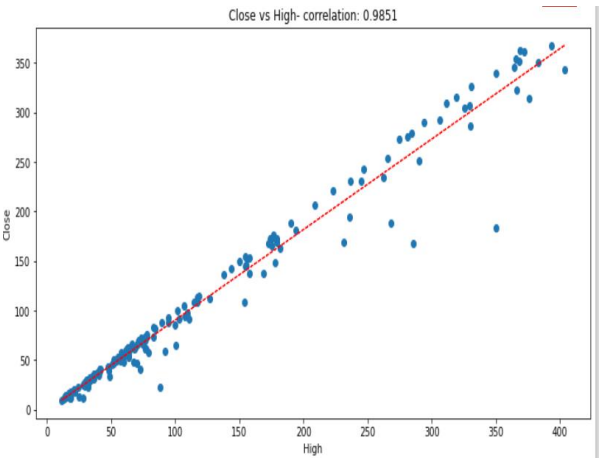




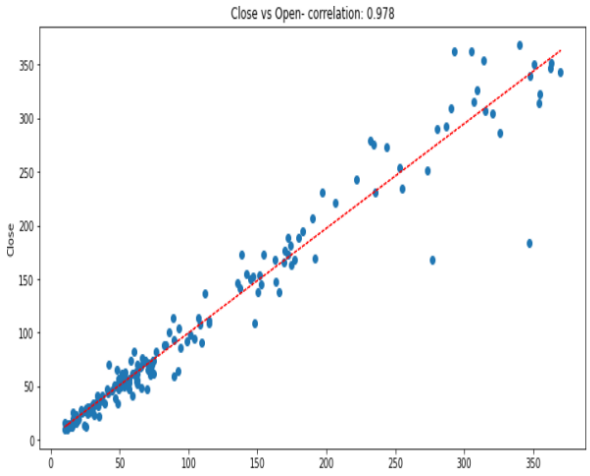
Plotting the independent variables. As we see , data is positively skewed, so we perform a log transform on it.We can see the transformed distribution which is similar to a normal distribution.

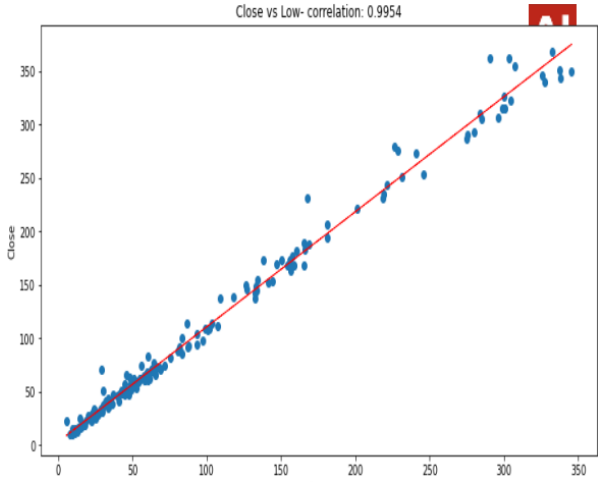






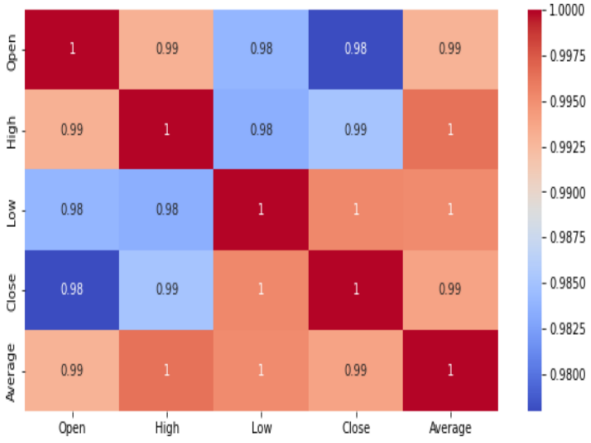
As we can see that there is linear relation and high correlation between each independent variables and our dependent variable.





There is a linear relation and very high correlation between our dependent variable and independent variables. The value of correlation between Close and Open is 0.978 and b/w Close and Low is 0.9954.

5. Heat Map



The correlation matrix helps us visualize the correlation of each parameter with respect to every other parameter. The heatmap on this slide that our dependent variable (close price) is highly correlated with all the other independent variables.

6. Model Implementation

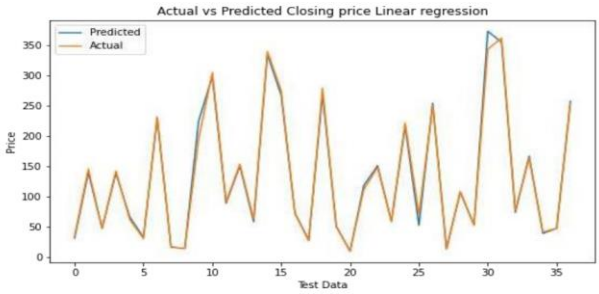
Following models have been implemented on our data:-

- Linear Regression
- Lasso Regression with Cross-validation
- Ridge Regression with Cross-validation
- Elastic Net Regression with Cross-validation

The performance of these models using various evaluation metrics such as:-

- Mean Absolute error.
- Mean squared error and RMSE
- R-squared and Adjusted R-squared

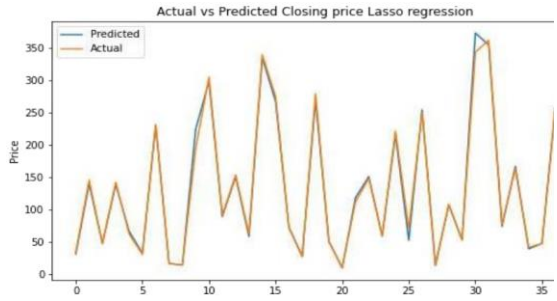
6.1 Linear regression



Simple Linear Regression Model predicted the closing price with Root Mean squared error (RMSE) of 8.3917

- R2 score of this model is 0.9937
- Adjusted R2 score has the value 0.9930 for this model. This tells us that around 99.3 percent of the variance in our dependent variable is attributable to the independent variables.

6.2 Lasso Regression with Cross-validation



Lasso Regression Model predicted the closing price with Root Mean squared error of 8.3864

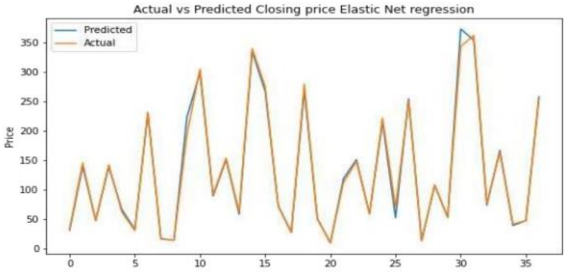
- R2 score of this model is 0.9938
- Adjusted R2 score has the value 0.9932 for this model. This tells us that around 99.32 percent of the variance in our dependent variable is attributable to the independent variables.

6.3 Ridge Regression with Cross-validation

Ridge Regression Model predicted the closing price with Root Mean squared error of 8.3824

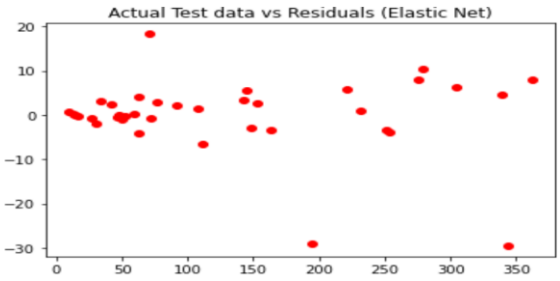
- R2 score of this model is 0.9938
- Adjusted R2 score has the value 0.9932 for this model. This tells us that around 99.32 percent of the variance in our dependent variable is attributable to the independent variables.

6.4 Elastic Net Regression with Cross-validation



Elastic Net Regression Model predicted the closing price with Root Mean squared error of 8.3760

- R2 score of this model is 0.9938
- Adjusted R2 score has the value 0.9932 for this model. This tells us that around 99.32 percent of the variance in our dependent variable is attributable to the independent variables.



The residuals (actual value – predicted) against the predicted values of our best performing model – Elastic Net regression. This is to check whether Heterodasceticity is present in our data or not. Since the data is symmetrical around zero, we can safely say that there is no heterodasceticity in our data. Hence the assumption of linear regression is valid here.

7. Evaluation Metrics

	Linear Regression	Ridge	Lasso	Elastic-Net
MAE	4.8168	4.8262	4.8334	4.8483
MSE	70.4204	70.3311	70.2641	70.1569
RMSE	8.3917	8.3864	8.3824	8.3760
R-square	0.9937	0.9938	0.9938	0.9938
Adjusted R-square	0.9930	0.9932	0.9932	0.9932

The best performing model is elastic net as it has higher accuracy and least error value

8. Conclusion:

- There is a high correlation between the dependent and independent variables. This is a good thing as we can make really accurate predictions using simple linear models.
- We implemented several models on our dataset in order to be able to predict the closing price and found that Elastic Net regressor is the best performing model with Adjusted R2 score value of 0.9932 and it scores well on all evaluation metrics.
- All of the models performed quite well on our data giving us the accuracy of over 99%.
- We found that there is a rather high correlation between our independent variables. This multicollinearity however is unavoidable here as the dataset is very small.
- We found that the distribution of all our variables is positively skewed. so we performed log transformation on them.
- Using data visualization on our target variable, we can clearly see the impact of 2018 fraud case involving Rana Kapoor as the stock prices decline dramatically during that period.
- With our model making predictions with such high accuracy even on unseen test data, we can confidently deploy this model for further predictive tasks using future real data.

References-

1. Stackoverflow
2. GeeksforGeeks
3. matplotlib.org