

Capstone Project - 2

Yes Bank Stock's Closing Price Prediction

By

Ishan Singh



A stock market is the aggregation of buyers and sellers of stock (also called shares), which represent ownership claims on businesses. The closing price is the raw price or cash value of the last transacted price in a security before the market officially closes for normal trading. It is often the reference point used by investors to compare a stock's performance since the previous day and closing prices

- Problem statement
- Data summary
- Feature engineering
- Exploratory Data Analysis
- Modelling approach
- Predictive modelling
 - Decision tree
 - Random forest
 - XG Boost
 - Prophet
- Model comparison
- Challenges
- Conclusion



Problem Statement

The challenge of this project is to accurately predict the future closing of a given stock across a given period of time in the future.



Data Summary

In the **YES BANK** stock closing dataset we have total 185 rows & 5 columns.

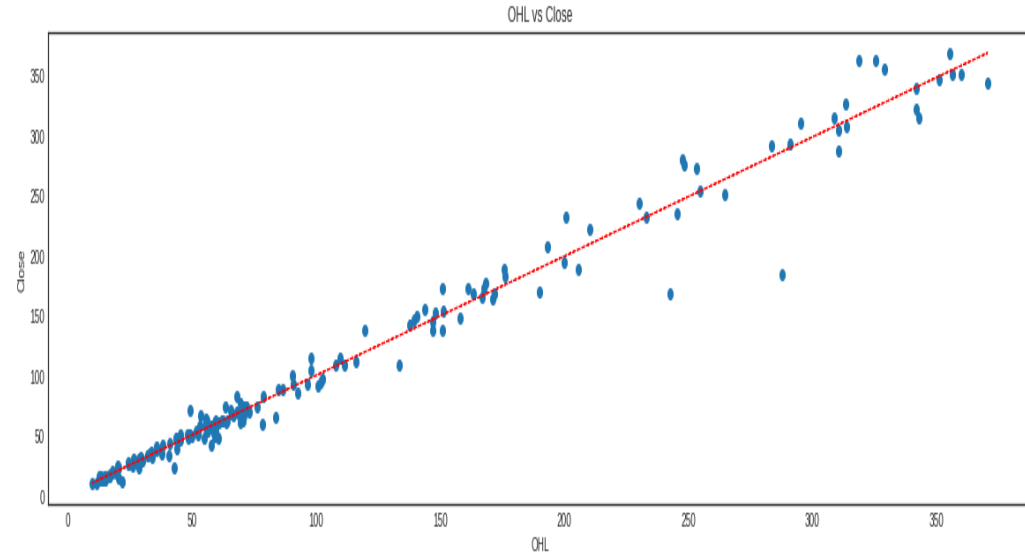
- Date : The Month and Year of the time the prices were recorded from 1st July 2005 To 1st November 2020.
- Open : Opening Stock price of the Month.
- Close : Closing Stock price of the Month.
- High : Highest stock price of the Month.
- Low : Lowest stock price of the Month.

Feature Engineering

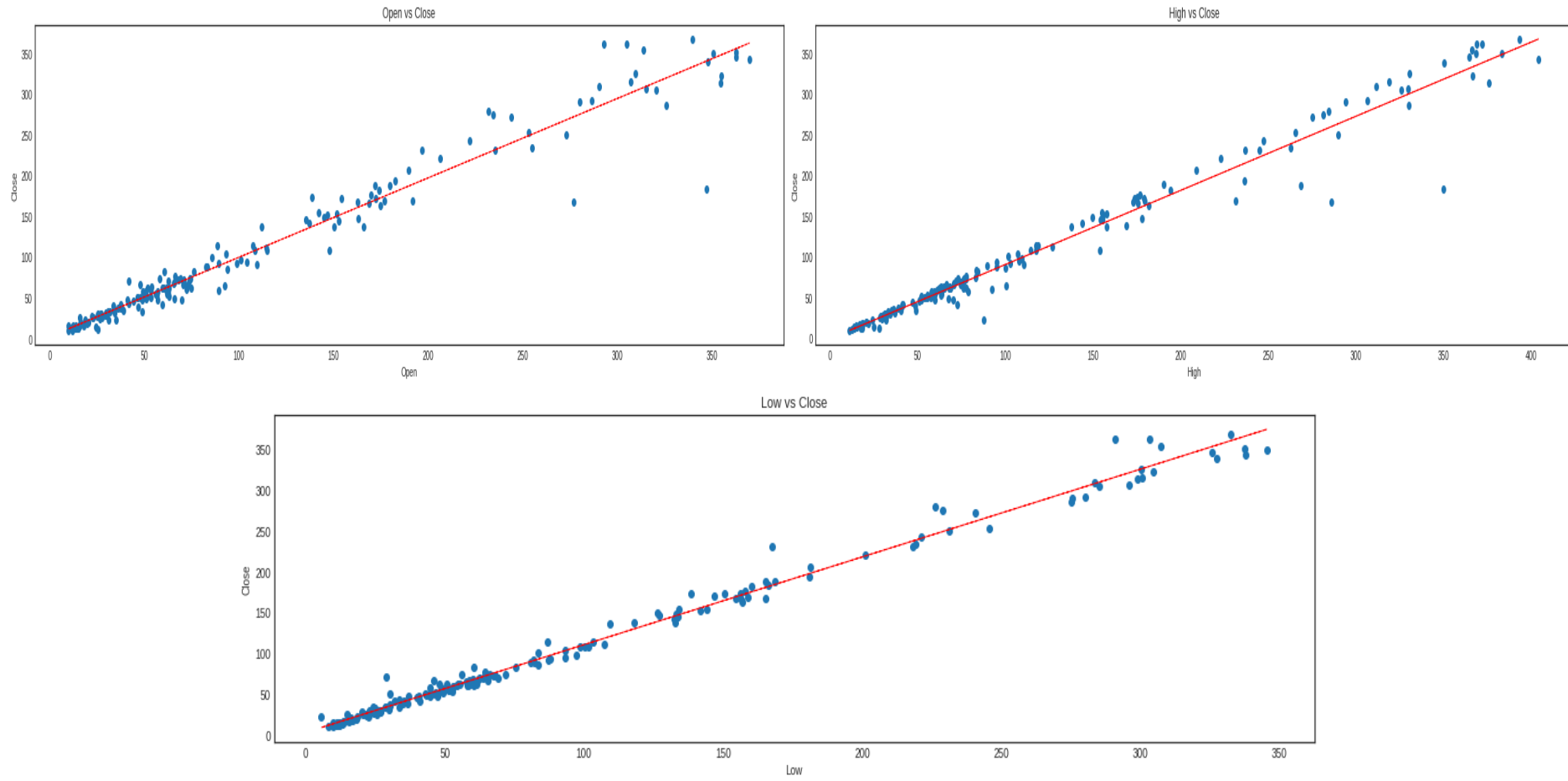
Linear Regression

Created a new feature based on average of other features in the given dataset using linear regression to check if the linear relationship exist between the target variable and independent variable through scatterplot.

- There are no missing values in the dataset



Feature Engineering

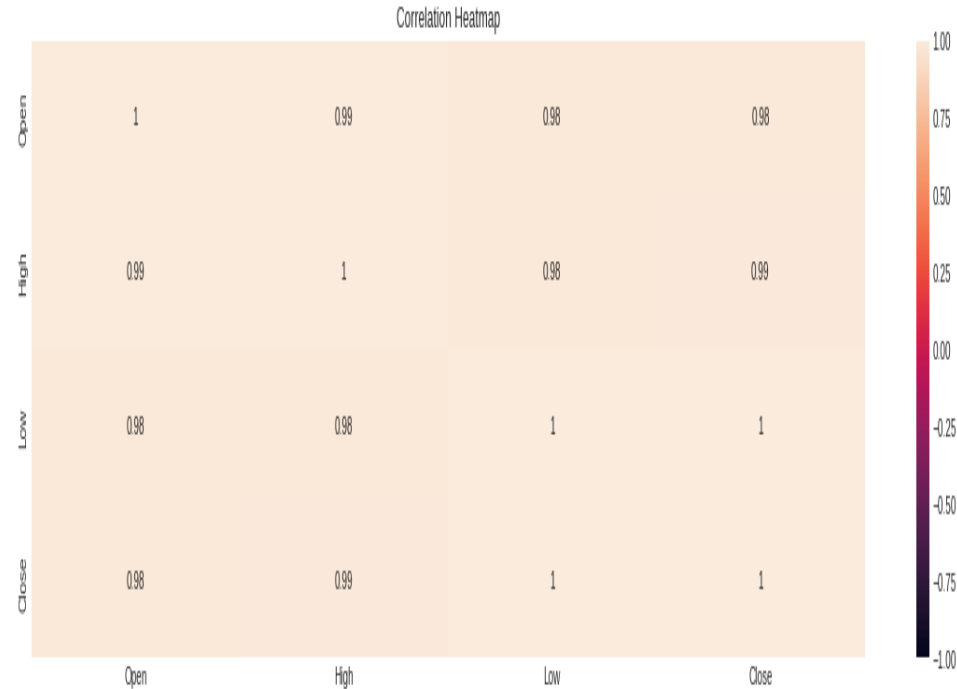


Exploratory Data Analysis(EDA)

Correlation Heatmap

we can see that every feature is extremely correlated with each other so we take just one feature or average of these features that would be sufficient for our regression model as linear regression assumes there is no multicollinearity in the features.

To reduce multicollinearity we use regularization that means to keep all the features but reducing the magnitude of the coefficients of the model. This is a good solution when each predictor contributes to predict the dependent variable.



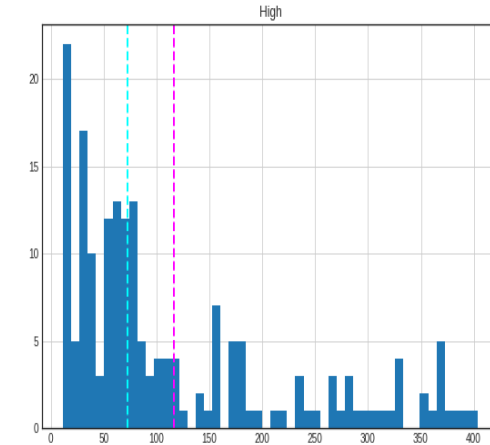
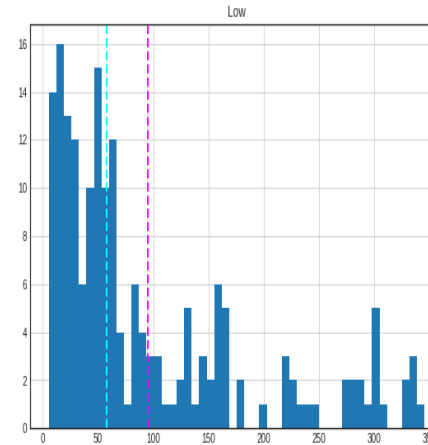
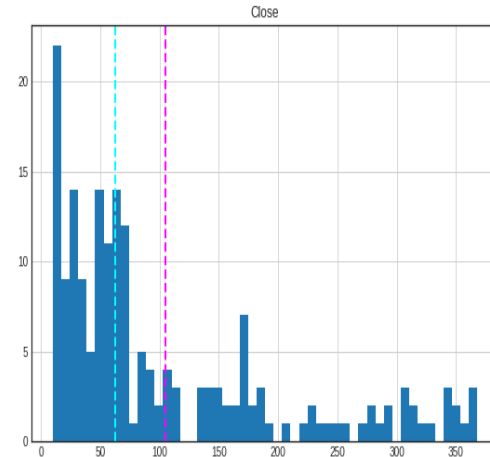
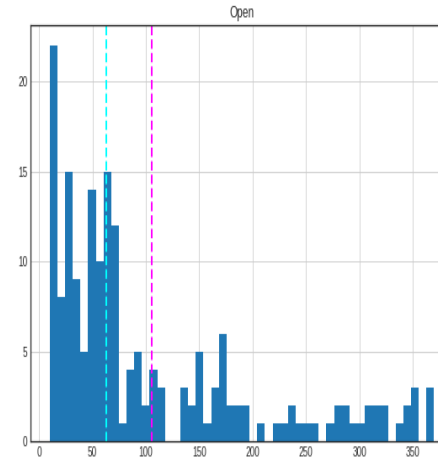
I have divided the stock data from 2005 to 2020 for Yes bank. Stock closing price boomed after February 2016 and peaked the highest closing price in July 2018. And the lowest closing price in March 2009.

- We can see that the lowest closing price in March 2009 is \$9.98.
- The highest closing price in July 2018 is \$367.9.

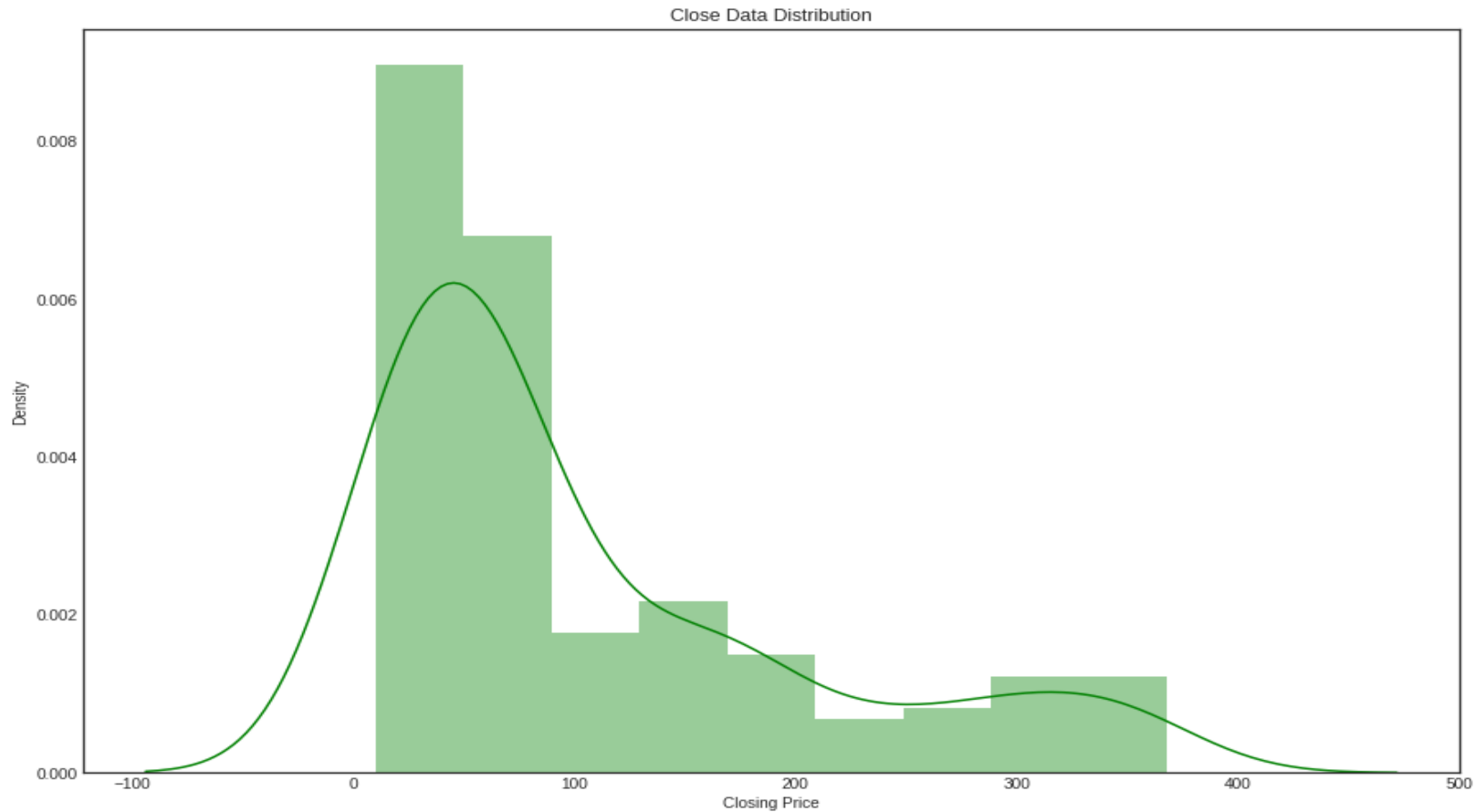


EDA (contd.)

As we can see ,the dependent variable is positively skewed.



EDA(contd.)



Modelling Approach

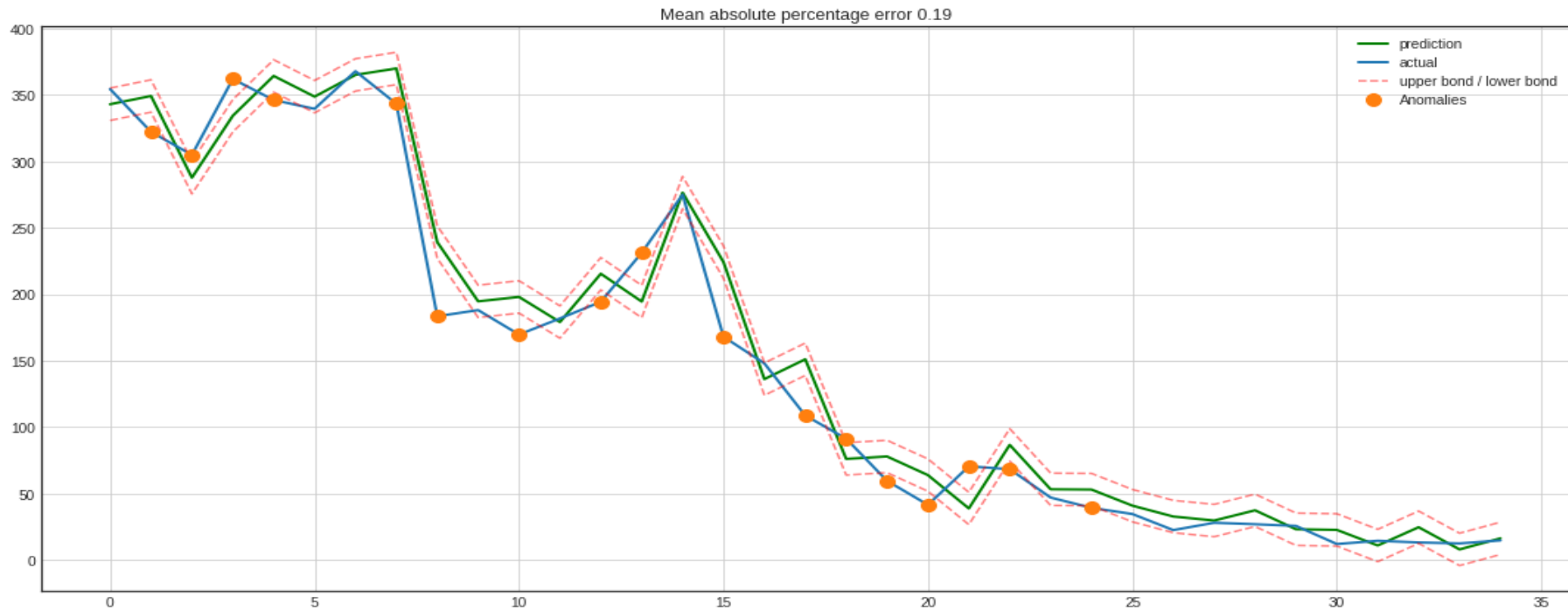
- Linear regression with regularizations
- Arima
- Prophet
- eXtreme Gradient Boost
- Random forest
- Decision Tree
- Choice of split is taken as K-fold cross validation with $k=5$, because of the computational power available and to reduce overfitting.
- Evaluation metrics is RMSE to punish outliers, and choose model that is able to generalize the results for all points including outliers.

Model Comparison

	model	mae	mse	rmse	R2_score	mape
1	linear regression	16.93	486.35	22.05	0.97	0.19
2	ridge regression	16.91	488.09	22.09	0.97	0.19
3	lasso regression	16.87	626.48	25.03	0.96	0.16
4	elastic net regression	18.34	726.63	26.96	0.95	0.16

Model Comparison

Ridge Regression



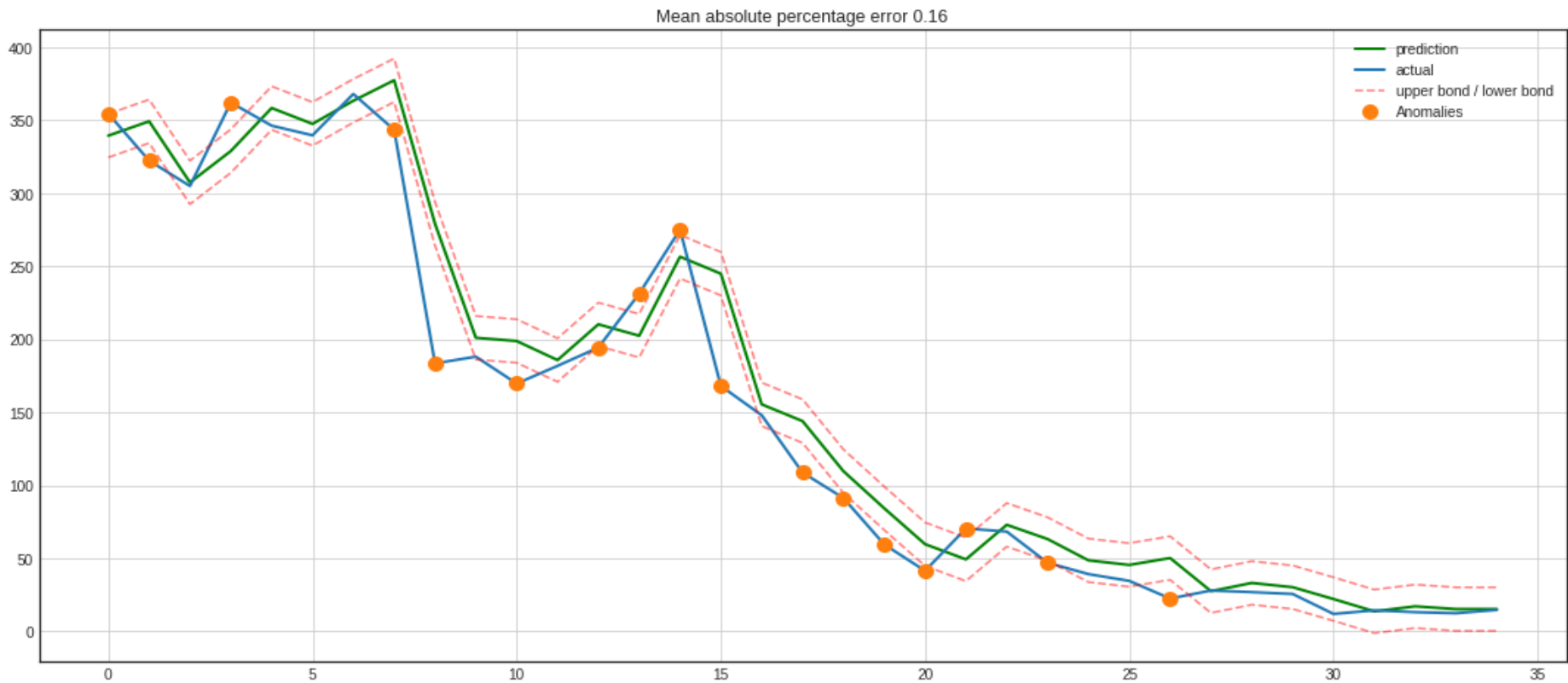
Model Comparison

Lasso Regression



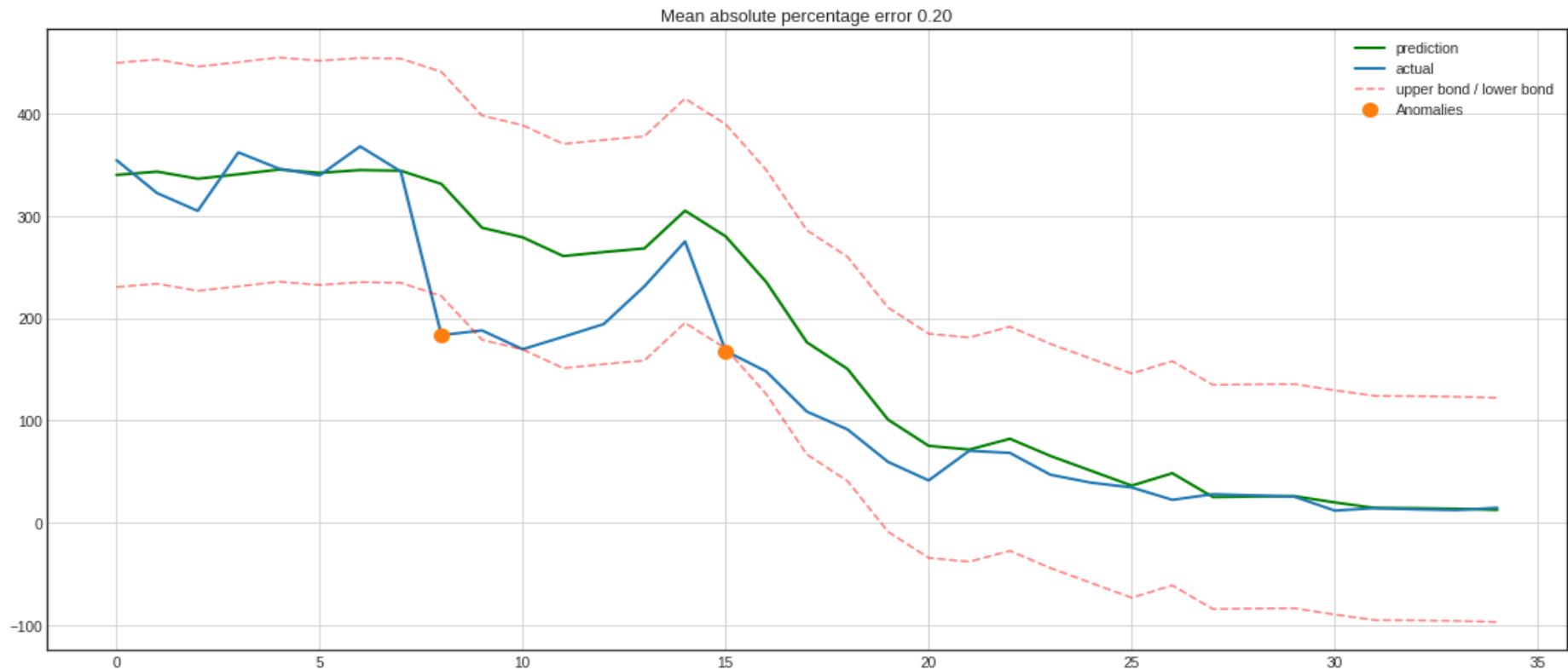
Model Comparison

Elasticnet Regression



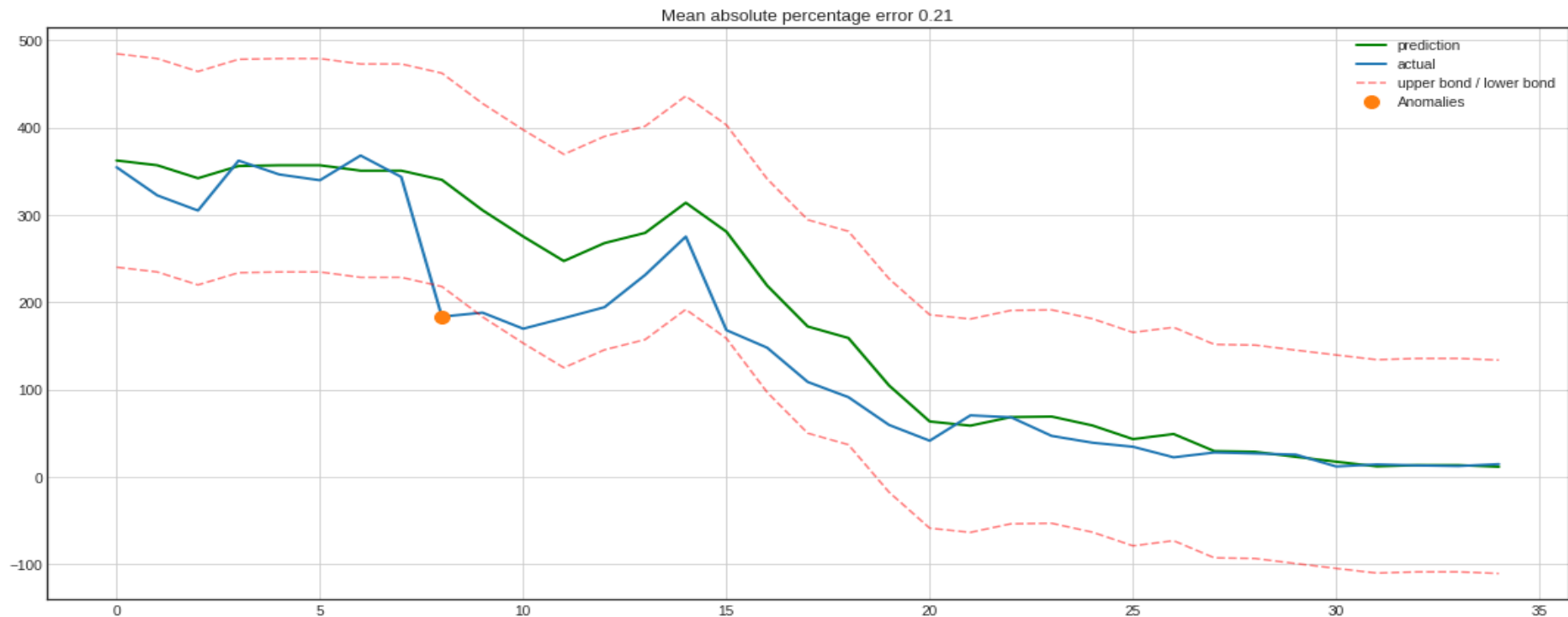
Model Comparison

Random Forest Regression



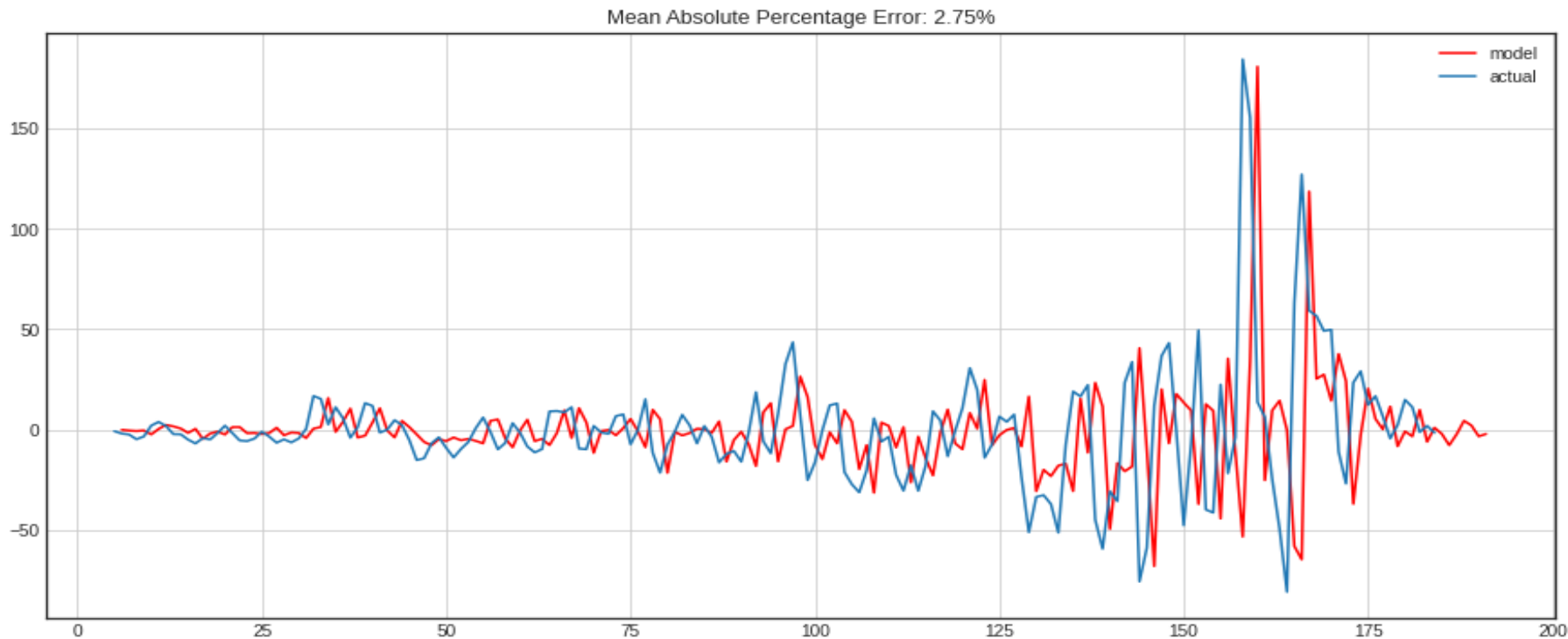
Model Comparison

eXtreme Gradient Boost



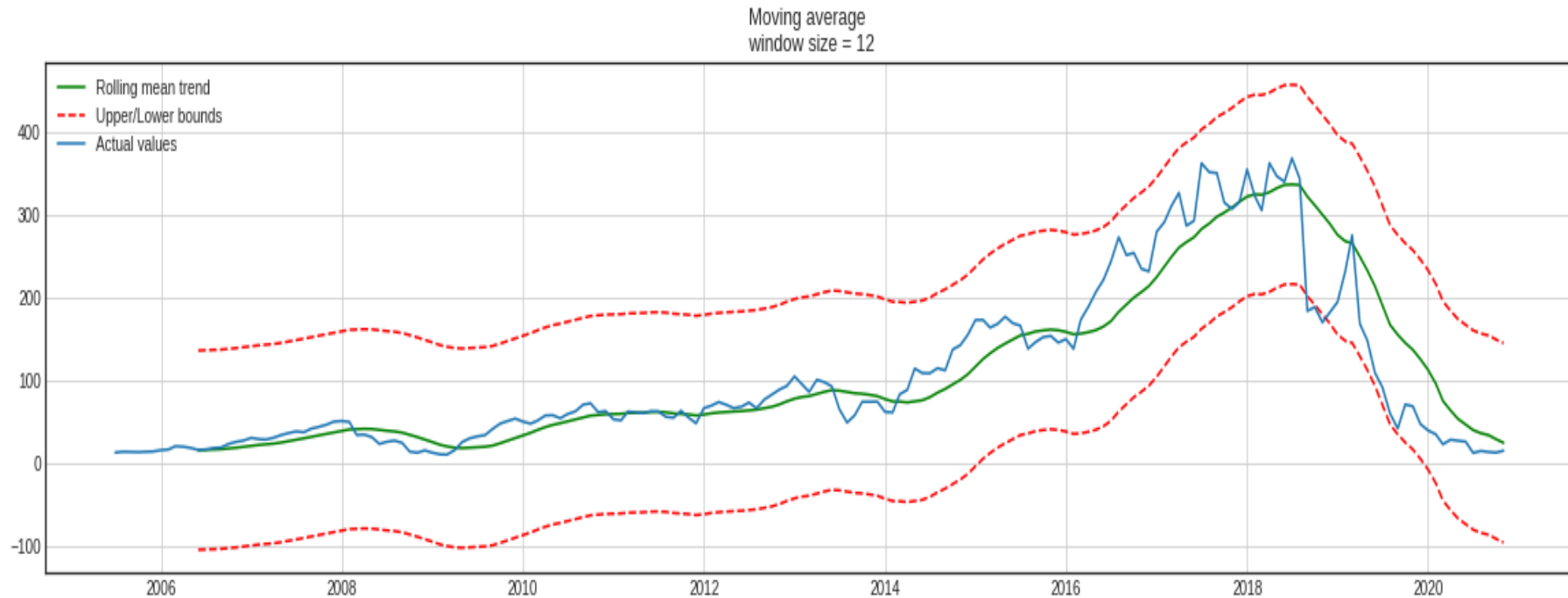
ARIMA MODEL

As we can see, the model can predict the trend of the actual stock closing prices very closely. The accuracy of the model can be enhanced by training with more data.



Moving Average

Prediction of stock's closing price for next month observed on previous 12 month data.



Challenges

- A huge amount of data needed to be dealt while doing the project which is quit an important task and also even small inferences need to be kept in mind.
- As dataset was quite big enough which led more computation time.



Conclusion

- We have successfully built a predictive models that can predict the stock closing price .
- Ridge regression model had the lowest RMSE.
- The proposed stock closing price predictor has been successfully trained by using regression model on the sample of datasets and the closing stock price prediction process has been successfully performed by using Prophet model being tested on the test dataset.

Thank You