

Capstone Project – 2

Supervised Machine Learning
-Regression

Yes Bank Stock Closing Price Prediction

Amol Rakhunde

Will Be Discussing On:

1. Problem Statement
2. Introduction
3. Data Cleaning
4. Exploratory Data Analysis (EDA)
5. Transforming Data
6. Splitting Data
7. Fitting Different Model
8. Cross Validation & Hyperparameter Tuning
9. Conclusion



1. Problem Statement:

- Prediction of Yes Bank stock closing price.
- Getting accuracy score of several machine learning model.
- Will be discovering important factors that govern hotel year to book a hotel room.



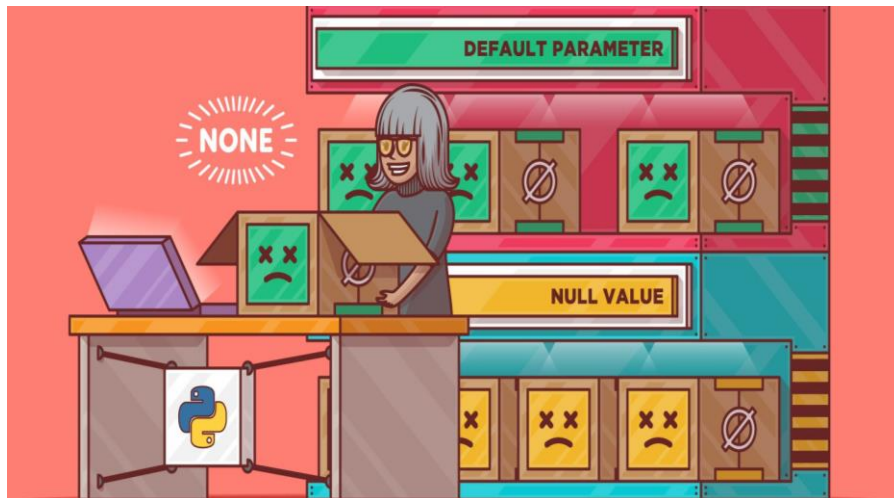
Problem Statements

2. Introduction:

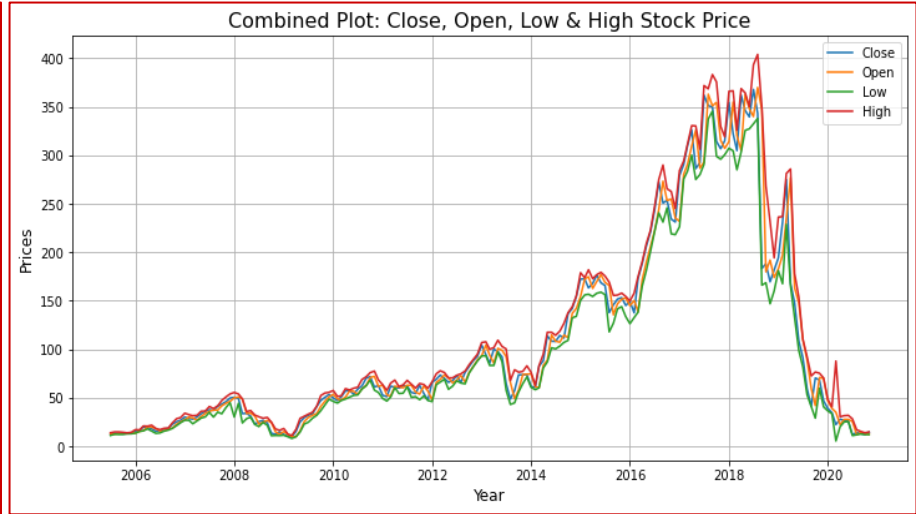
- **Date:** Monthly observation of stock prices since its inception.
- **Open:** The price of a stock when stock exchange market open for the day.
- **Close:** The price of a stock when stock exchange market closed for the day.
- **High:** The maximum price of a stock attained during given period of time.
- **Low:** The minimum price of a stock attained during given period of time.

3. Data Cleaning

- Null Values Treatment
- Duplicated Values Treatment
- Date Format Change
(i.e from Jul-05 to 2005-07-01)



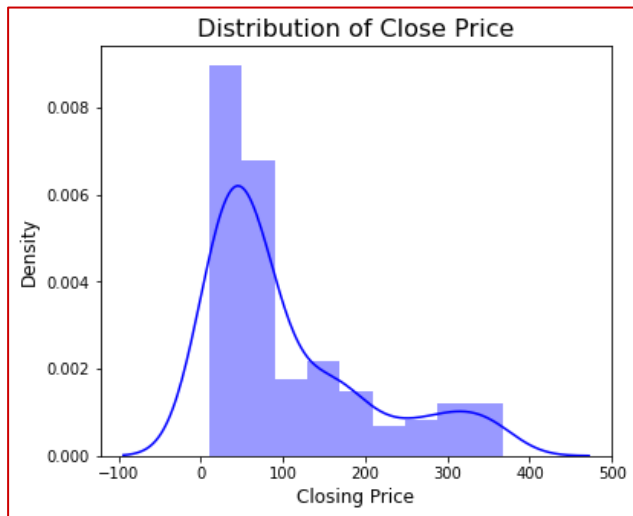
4. Exploratory Data Analysis (EDA)



- Sudden fall in stock after 2018 which justify the effect of fraud case involving Rana Kapoor.
- From combined plot, it shows strong correlation between each feature.

EDA (Continued)

- Distribution of Closing Price:



- Distribution of closing price is right skewed.
- We need this distribution to be normal distribution for training algorithm.

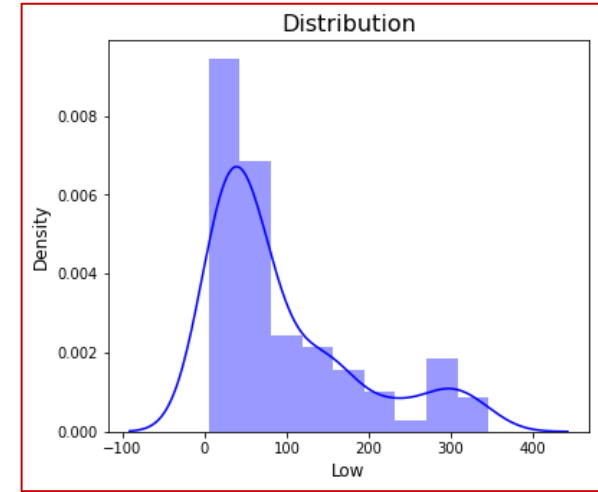
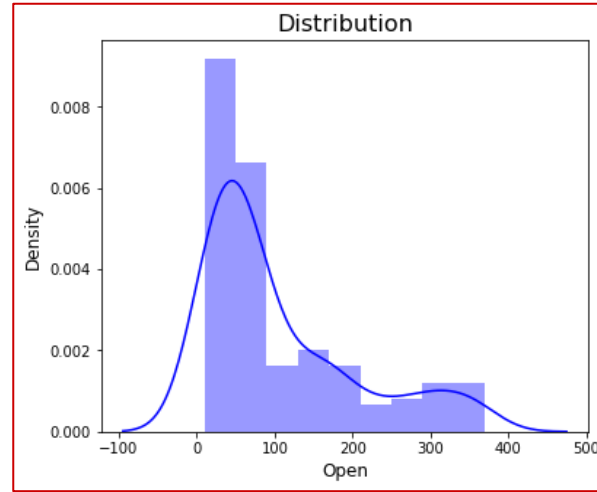
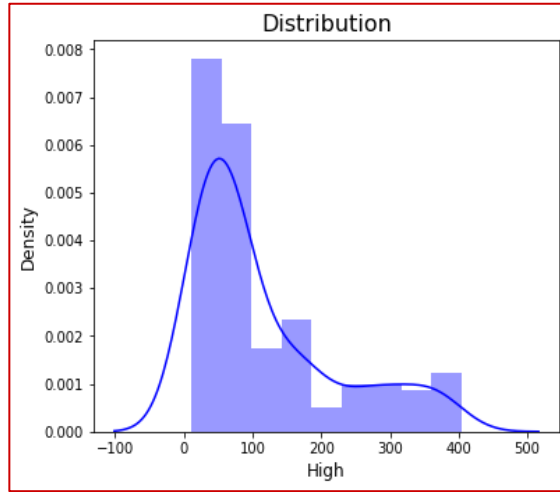
- After Log Transformation



- Distribution of closing price is normal distribution.

EDA (Continued)

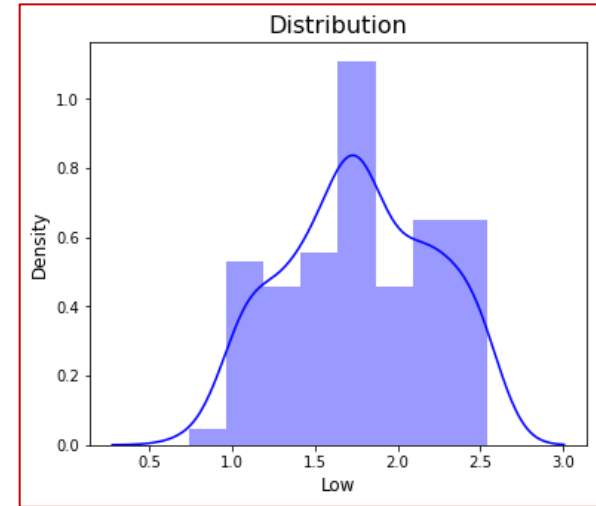
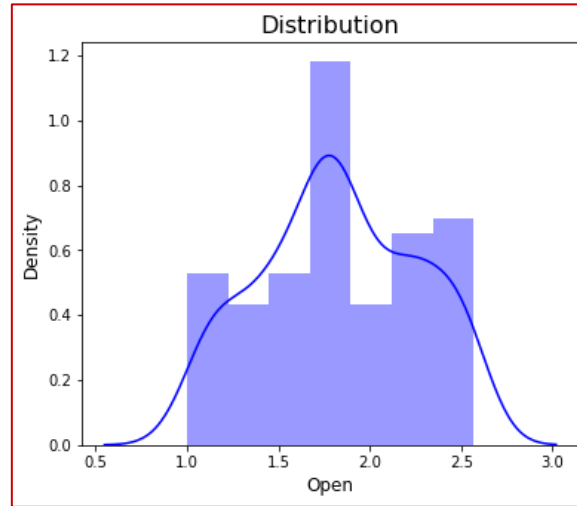
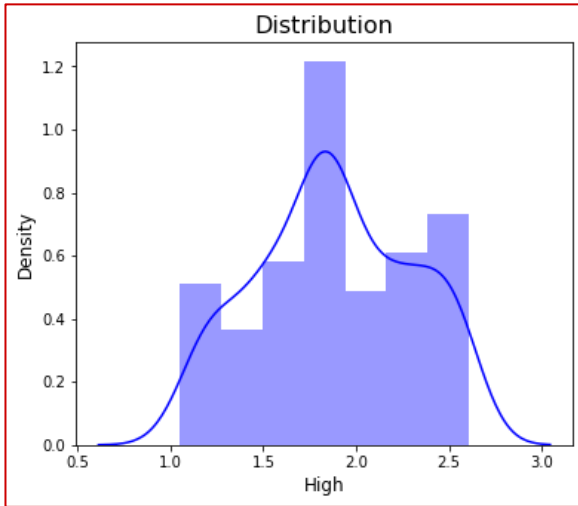
- Distribution of Open, High & Low Price of a stock:



- Distribution of opening price, high price and low price are also right skewed.
- Log transformation applied to make this distribution normal.

EDA (Continued)

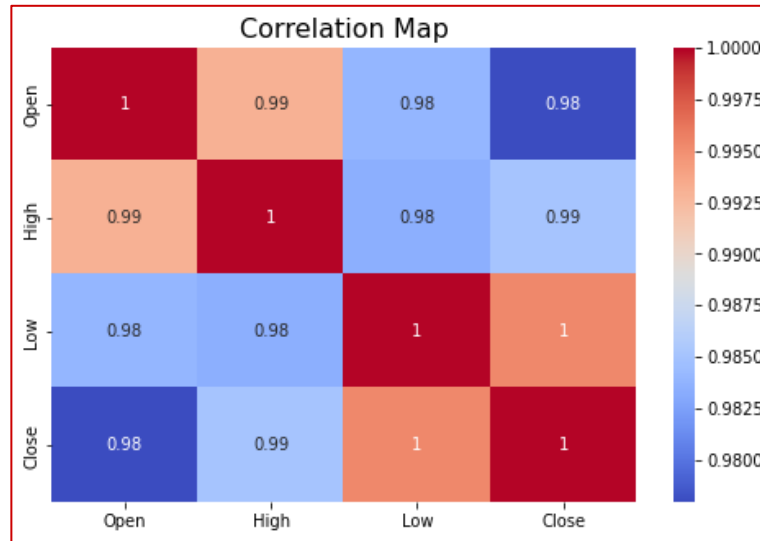
- Distribution of Open, High & Low Price of a stock after Log Transformation:



- Distribution of opening price, high price and low price are now normal distribution.

EDA (Continued)

- Correlation:



- All the features are strongly correlated with each other.

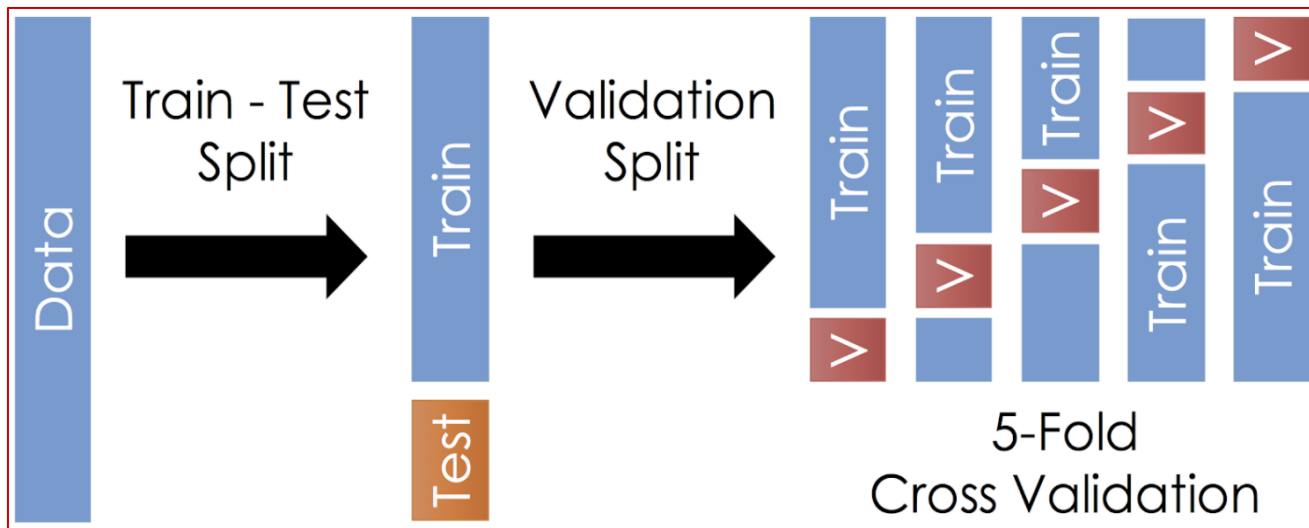
5. Transformation of Data

- To scale data into a uniform format that would allow us to utilize the data in a better way.
- For performing fitting and applying different algorithms to it.
- The basic goal was to enforce a level of consistency or uniformity to dataset.



6. Splitting Data

- Data splits into training dataset and testing dataset.
- Training dataset is for making algorithm learn and train model.
- Test dataset is for testing the performance of train model.
- Here 80% of data taken as training dataset & remaining 20% of dataset used for testing purpose.

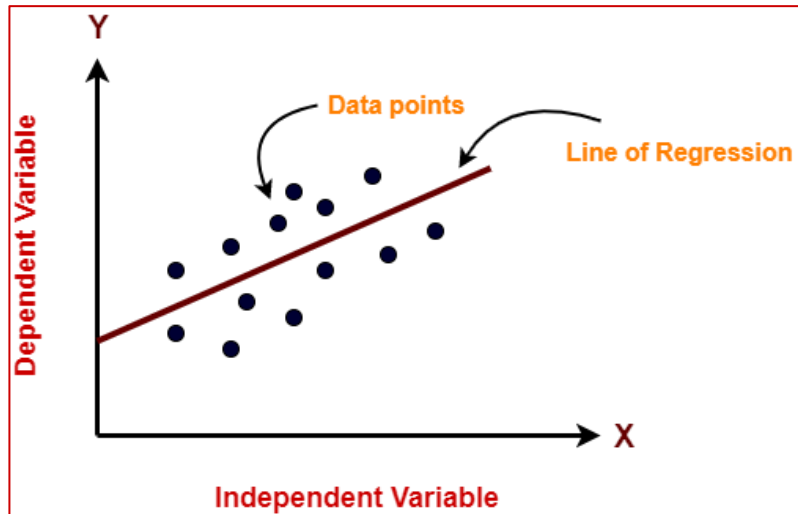


7. Fitting Different Model

7.1 Linear Regression

- Linear regression is one of the easiest and most popular Machine Learning algorithms.
- It is a statistical method that is used for predictive analysis.
- Linear regression algorithm shows a linear relationship between a dependent and independent variable; hence it is called as linear regression.

Evaluation Metrics: Linear Regression				
MSE	RMSE	MAE	MAPE	R2
0.032	0.178	0.151	0.095	0.823



7.2 Lasso Regression

- Lasso: Least Absolute Shrinkage and Selection operator
- It is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model.
- This method performs L1 regularization.

Evaluation Metrics: Lasso Regression				
MSE	RMSE	MAE	MAPE	R2
0.032	0.179	0.152	0.096	0.820

7.3 Ridge Regression

- Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity.
- When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.
- This method performs L2 regularization.

Evaluation Metrics: Ridge Regression				
MSE	RMSE	MAE	MAPE	R2
0.032	0.178	0.151	0.095	0.823

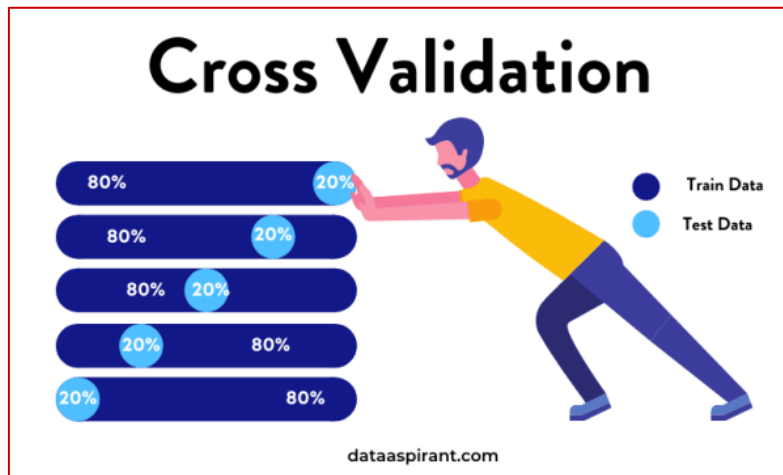
7.2 Elastic Net

- Elastic net is a popular type of regularized linear regression that combines two popular penalties, specifically the L1 and L2 penalty functions.
- Elastic Net is an extension of linear regression that adds regularization penalties to the loss function during training.

Evaluation Metrics: Elastic Net				
MSE	RMSE	MAE	MAPE	R2
0.036	0.191	0.157	0.102	0.820

8. Cross Validation & Hyperparameter Tuning

- It is a resampling procedure used to evaluate machine learning models on a limited data sample.
- Basically, Cross Validation is a technique using which Model is evaluated on the dataset on which it is not trained that is it can be a test data or can be another set as per availability or feasibility.
- Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting.

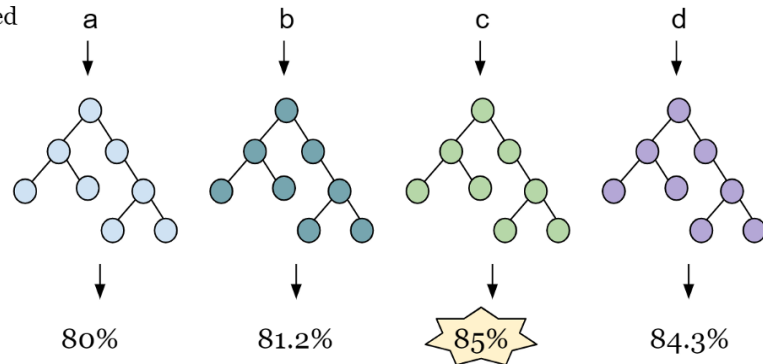


Test_Hyperparameters = [a, b, c, d]

Hyperparameter used
to make the model

model

Accuracy on test set



- Cross Validation & Hyperparameter tuning on Lasso Regression

Evaluation Metrics: CV & tuning on Lasso Regression				
MSE	RMSE	MAE	MAPE	R2
0.032	0.180	0.153	0.097	0.819

- Cross Validation & Hyperparameter tuning on Ridge Regression

Evaluation Metrics: CV & tuning on Ridge Regression				
MSE	RMSE	MAE	MAPE	R2
0.033	0.180	0.153	0.097	0.817

- Cross Validation & Hyperparameter tuning on Elastic Net

Evaluation Metrics: CV & tuning on Elastic Net				
MSE	RMSE	MAE	MAPE	R2
0.032	0.180	0.153	0.097	0.819

9. Evaluation Metrics Comparison

	Model	MSE	RMSE	MAE	MAPE	R2
0	LinearRegression	0.032	0.178	0.151	0.095	0.823
1	Ridge	0.032	0.178	0.151	0.095	0.823
2	Lasso	0.032	0.179	0.152	0.096	0.820
3	ElasticNet	0.036	0.191	0.157	0.102	0.796

10. Conclusion

- 1. We got a maximum accuracy score of 82%.
- 2. Linear, lasso and ridge regression show almost same R squared values.
- 3. Whereas elastic net model shows lowest R squared value and high MSE, RMSE, MAE & MAPE values.
- 4. Close, Open and high price of stock are strongly correlated with each other.