CAPSTONE PROJECT- 02

Supervised ML-Regression

Yes Bank Stock Closing Price Prediction

RANJIT GHAGDE

POINTS OF DISCUSSION:

1. Problem Statement
2. Introduction
3. Sample of Data
4. Data Cleaning
5. Heat Map Of Null Values
6. Linear Correlation
7. Exploratory Data Analysis (EDA)
8. Correlation Map
9. Splitting Data
10. Fitting Different Model
11. Cross Validation & Hyperparameter Tuning
12. Conclusion

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

In short:

- Prediction of Yes Bank stock closing price.
- Getting accuracy score of several machine learning model.

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. SAMPLE OF DATA

	Date	Open	High	Low	Close
0	Jul-05	13.00	14.00	11.25	12.46
1	Aug-05	12.58	14.88	12.55	13.42
2	Sep-05	13.48	14.87	12.27	13.30
3	Oct-05	13.20	14.47	12.40	12.99
4	Nov-05	13.35	13.88	12.88	13.41

1. The given dataset consist of 185 rows and 5 columns

2. The stock prices in this dataset are from the year 2005 to 2021

4. DATA CLEANING

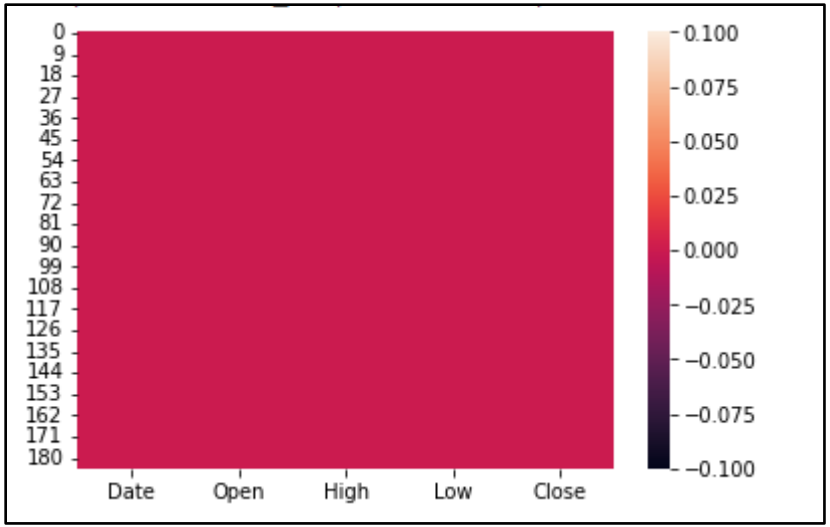
- Duplicated Values Treatment
- Date Format Change
- Null Values Treatment


```
[ ]  df.isna().sum()

Date      0
Open      0
High      0
Low       0
Close     0
dtype: int64
```

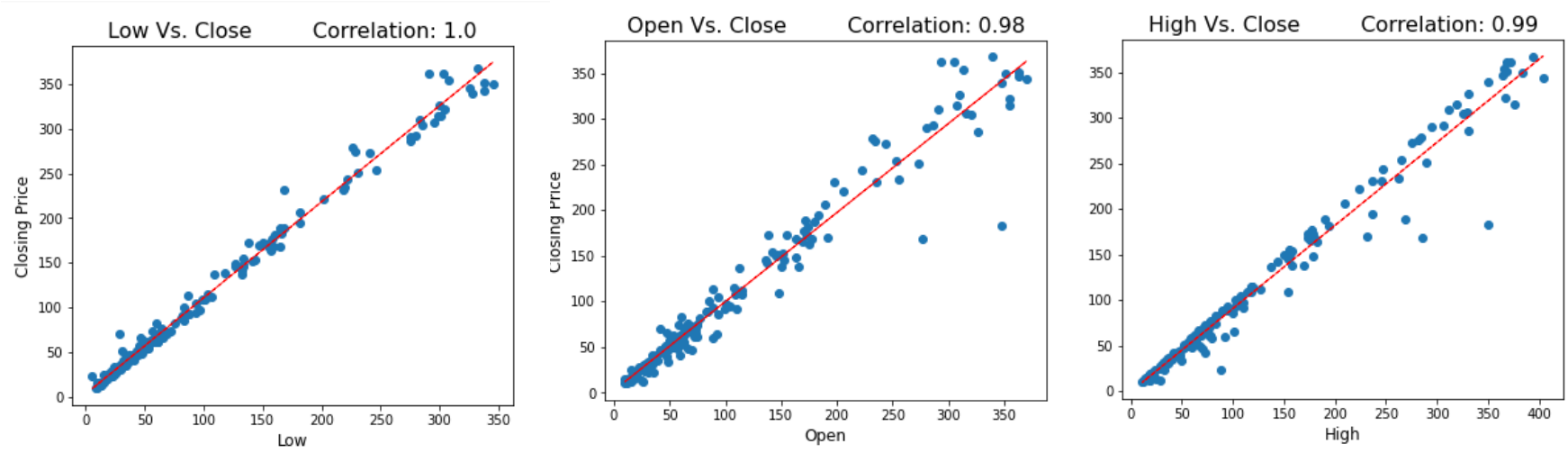
- Hence no null value present in data

5. HEATMAP



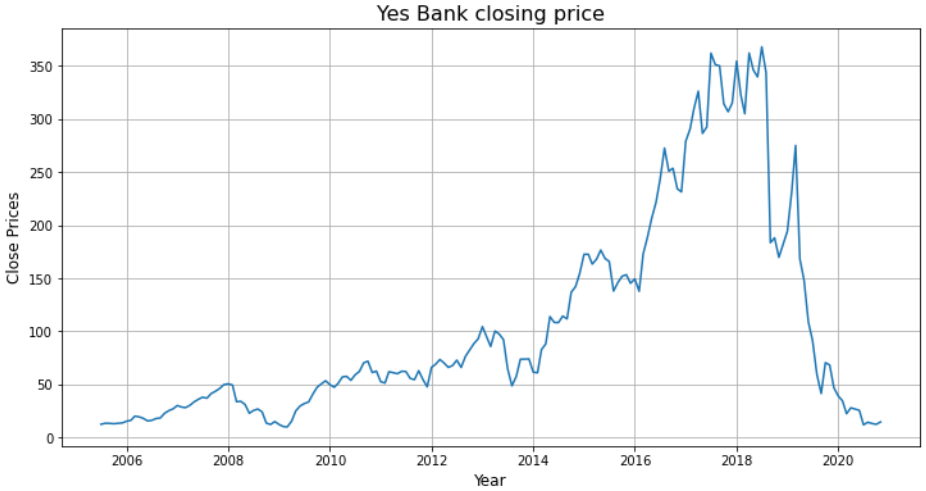
- We can see no null value is present in data.

6. LINEAR CORRERATION



- Above graph clearly shows that all features are linearly correlated.

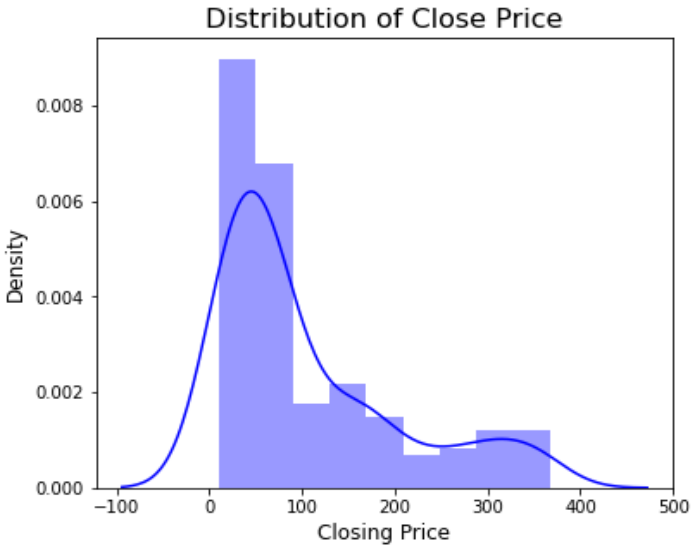
7. EDA: Exploratory Data Analysis



- 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.

7. EDA: Exploratory Data Analysis

➢ Distribution of Closing



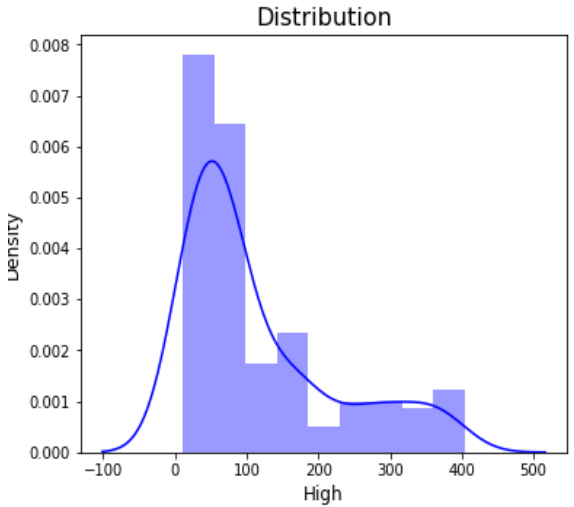
- Distribution of closing price is right skewed.
- We need this distribution to be normal distribution for training algorithm.

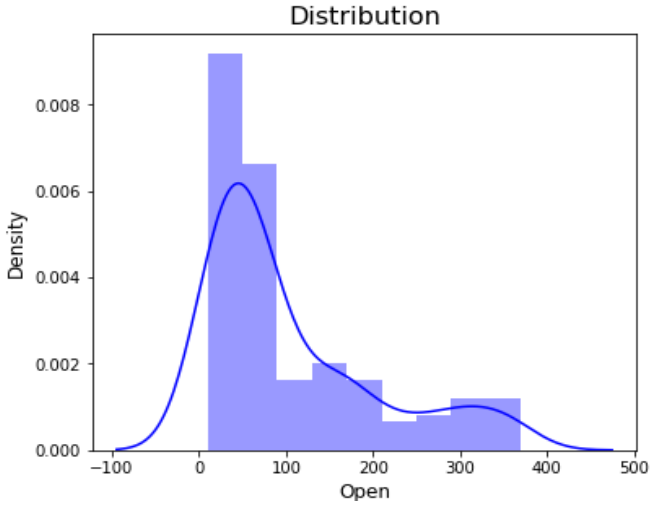
➢ After Log Transformation

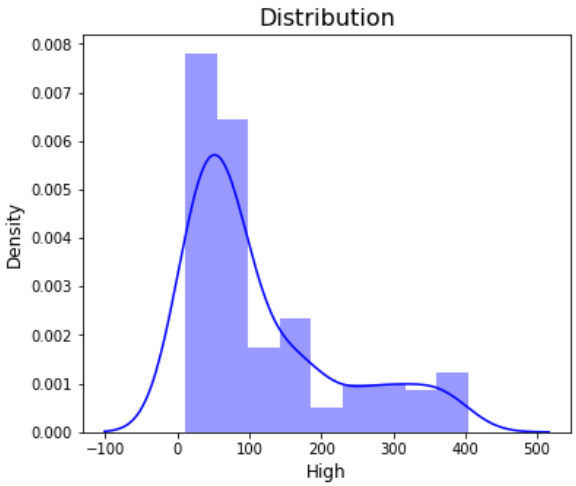
- Normal Distribution ofclosing price.

7. EDA: Exploratory Data Analysis

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



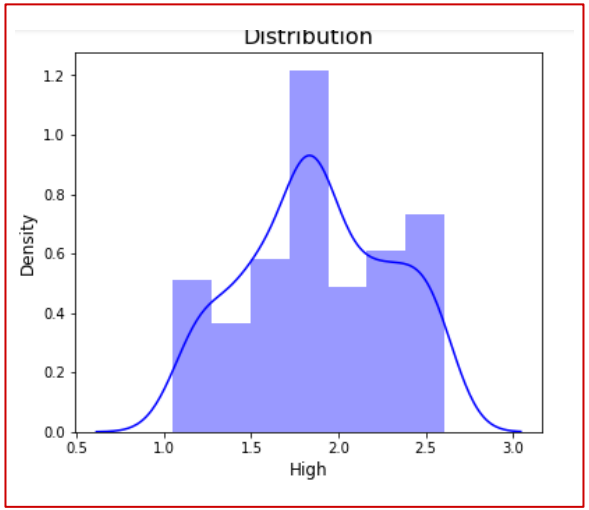


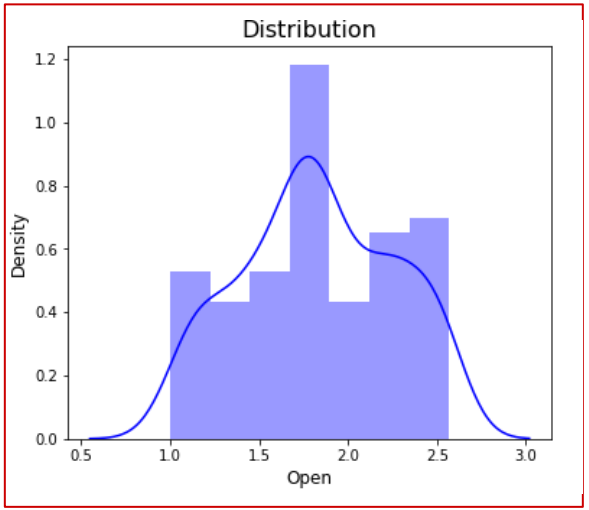


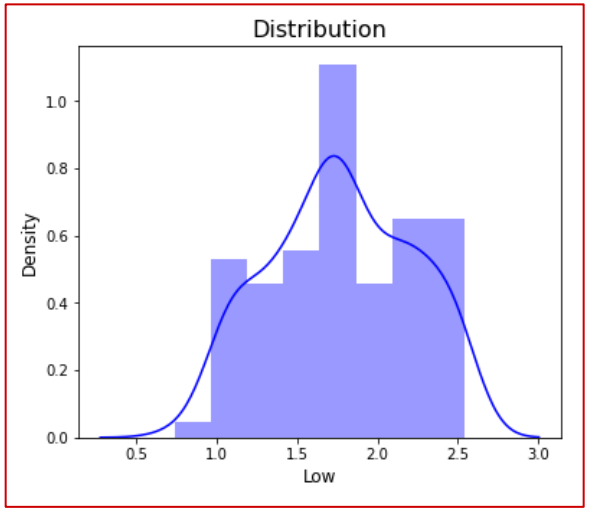
- Distribution of opening price, high price and low price are also right skewed.
- Log transformation applied to make this distribution normal.

7. EDA: Exploratory Data Analysis

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



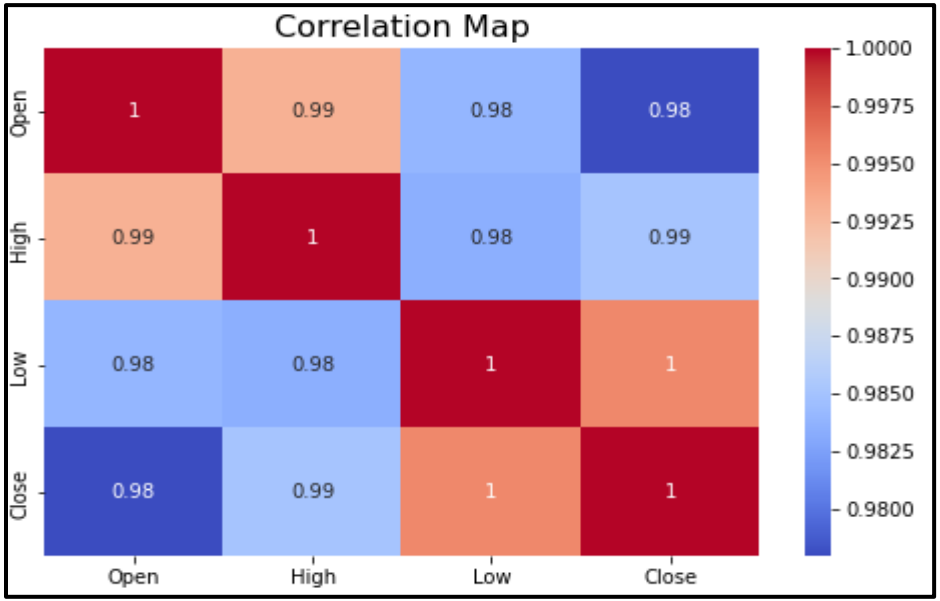




- Distribution of opening price, high price and low price are now normal distribution.

8. CORRELATION MAP

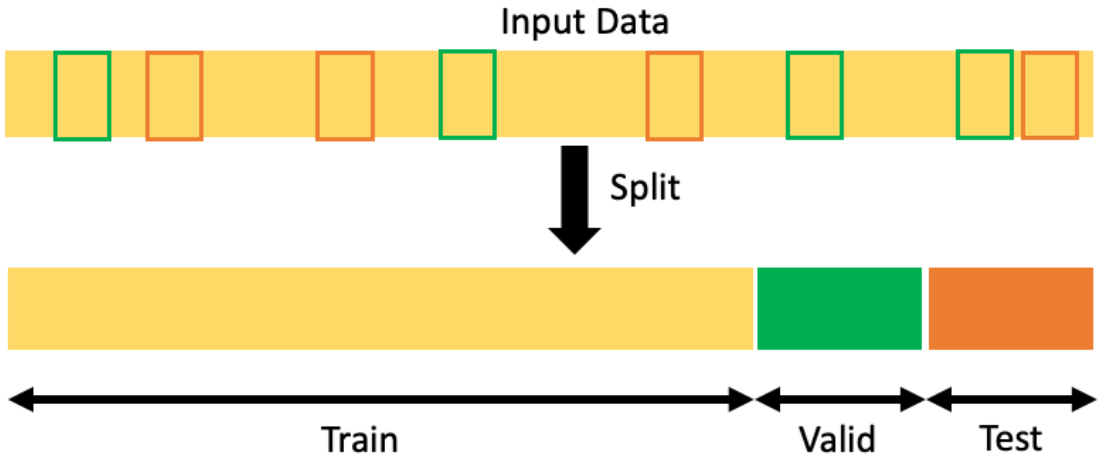
➢ Correlation:



- All the features are strongly correlated with each other.

9. SPLITING 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.



10. FITIING DIFFERENT MODEL

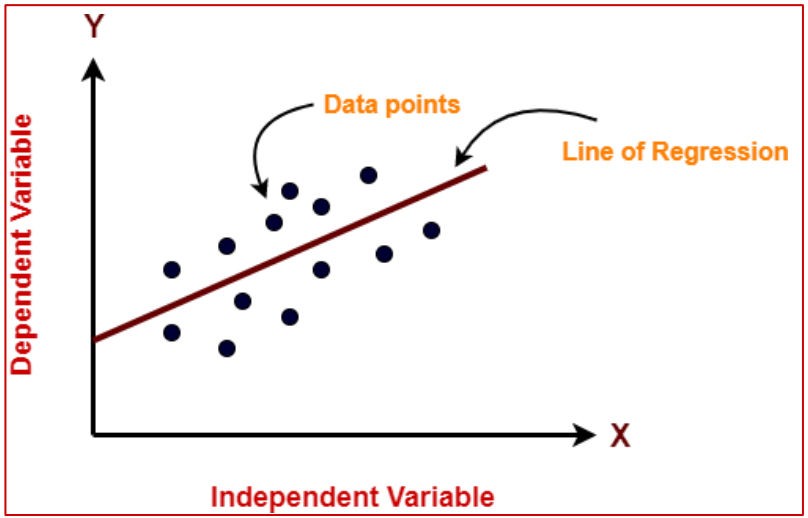
A] Linear Regression

• 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
0.167	0.409	0.348	0.095	0.823



10. FITIING DIFFERENT MODEL

B] 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
0.169	0.412	0.35	0.096	0.82

10. FITIING DIFFERENT MODEL

C] Ridge Regression

- Ridge regression is a model tuning method that is used to analyses 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
0.167	0.409	0.348	0.095	0.823

10. FITIING DIFFERENT MODEL

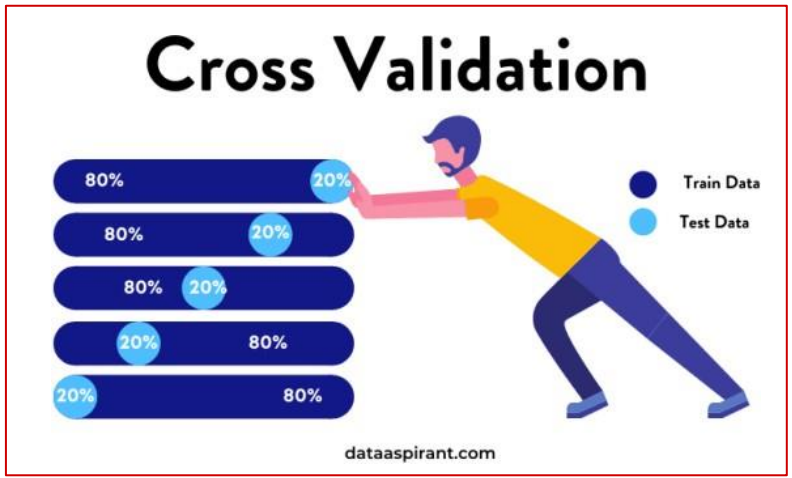
D] Elastic Net

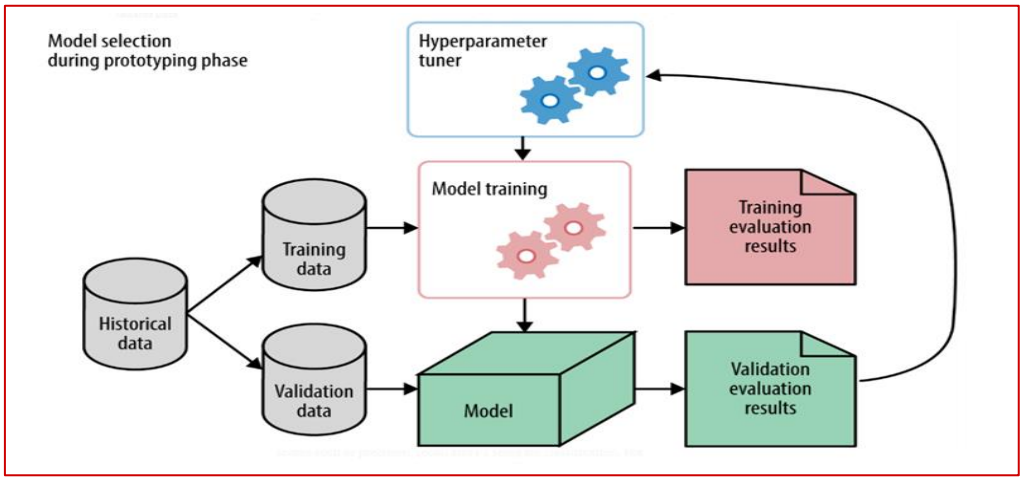
- Elastic net is a popular type of regularized linear regression that combines two popular penalties,
- specifically the L1 and L2 penalty functions.
- Elastic Net is an extension of linear regression that adds regularization penalties to the loss function during training.

Evaluation Metrics: Elastic Net				
MSE	RMSE	MAE	MAPE	R2
0.0178	0.422	0.357	0.099	0.811

11. CROSS VALIDATION & HYP. 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.





- **Cross Validation & Hyperparameter tuning on Lasso Regression**

Evaluation Metrics: CV & tuning on Lasso Regression				
MSE	RMSE	MAE	MAPE	R2
0.17	0.412	0.351	0.096	0.82

- **Cross Validation & Hyperparameter tuning on Ridge Regression**

Evaluation Metrics: CV & tuning on Ridge Regression				
MSE	RMSE	MAE	MAPE	R2
0.0173	0.415	0.352	0.097	0.817

- **Cross Validation & Hyperparameter tuning on Elastic Net**

Evaluation Metrics: CV & tuning on Elastic Net				
MSE	RMSE	MAE	MAPE	R2
0.017	0.412	0.351	0.096	0.82

12. EVALUTION MATRIX COMPARISON

	Model	MSE	RMSE	MAE	MAPE	R2
0	LinearRegression	0.167	0.409	0.348	0.095	0.823
1	Ridge	0.167	0.409	0.348	0.095	0.823
2	Lasso	0.169	0.412	0.350	0.096	0.820
3	ElasticNet	0.178	0.422	0.357	0.099	0.811

13. 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.