

YES BANK STOCK CLOSING PRICE PREDICTION

Amol Kale, Mukund Pandit,
Ayushi Jain, Siddharth Budhiraja
Data science trainees, AlmaBetter

Abstract:

Yes Bank is a well-known bank in the Indian financial domain. Yes Bank is listed on exchange market. Stock price of any stock may depend on several factors. Incidents such fraud case involving bank management people, certainly have huge impact on stock price. Such fraud case involving Rana Kapoor happened in Yes Bank in year 2018. After which stock price of Yeas Bank went down severely.

Investment or trading in a stock market, it may be highly rewarding or it may be highly risky. So, to reduce the risk on investment and increase the profitability of investment in a stock, prediction about stock closing price could be very useful. Stock prices such opening price, closing price, high price and low-price during period of time, may lead us to predict future closing price of that stock.

Our experiments can help understand and predict about the closing price of a stock. Data analysis and prediction using machine leaning algorithm leads to feature selection and predict future closing price of a stock.

Keywords: *machine learning, yes bank stock prediction, linear regression, lasso, ridge, elastic net, EDA.*

1.Problem Statement

Data is about the Yes Bank stock prices for each month from its inception. Data provided contains opening price, closing price, high price and low-price during each month. These features altogether will give past trend of closing prices of a stock. Lot many factors will have impact on closing price of a stock.

The main objective is to build a predictive model, which could help them in predicting the closing price of Yes Bank stock proactively. This would in turn help them in predicting closing price a stock and do investment and trading in that particular stock.

2. Introduction

Share market where exchange of stocks of several companies takes place. On every weekday excluding exchange holidays market will open at 9:00am and close at 3:30pm. Investors and traders do their trades on stock market during this hour. These investment comes with huge risks. And these investments can have huge profit also.

So, there is need of minimizing the risk on investment and increase its profitability. Learning the past trends of stock and to be able to predict the closing price of a stock

which leads to take better investment decisions. Closing price of stocks depends on lot many factors. Dataset contains opening price, closing price, high price and low-price for each month. Details of these are as follows:

- **Date:**
In our data its monthly observation of stock since its inception.
- **Open:**
The price a stock when the stock exchange open for the day.
- **Close:**
The price of a stock when the stock exchange 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. Steps involved:

- **Null values Treatment**
Null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result. Our dataset dose not contains any null values.
- **Duplicate Values**
Duplicate values dose not contribute anything to accuracy of results. Large duplicate may lead to slower computation and higher space requirement. Our dataset dose not contains any duplicate values.

- **Cleaning Data**

Features in dataset must be of required data type, which makes it easy for handling. Date feature is in object type data. So, by parsing this string date into datetime data type. Format of date changed to YYYY-MM-DD from MMM-YY.

- **Exploratory Data Analysis**

Closing price is out dependent variable. Plot of monthly closing price of Yes Bank stock from inception shows the performance of that stock.

From above plot, there is sudden fall in stock after 2018 which justify the



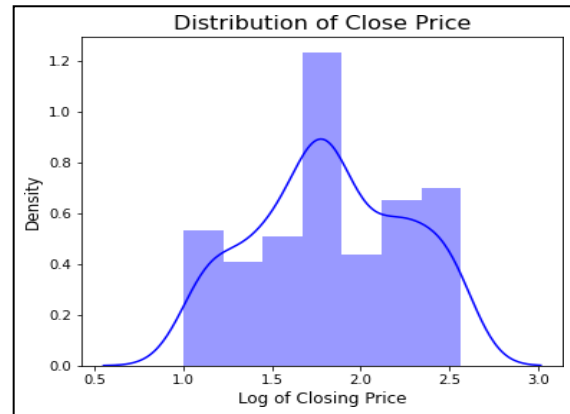
effect of fraud case involving Rana Kapoor.

Next, we combine the plots of each feature that is open, close, high and low price of stock. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.



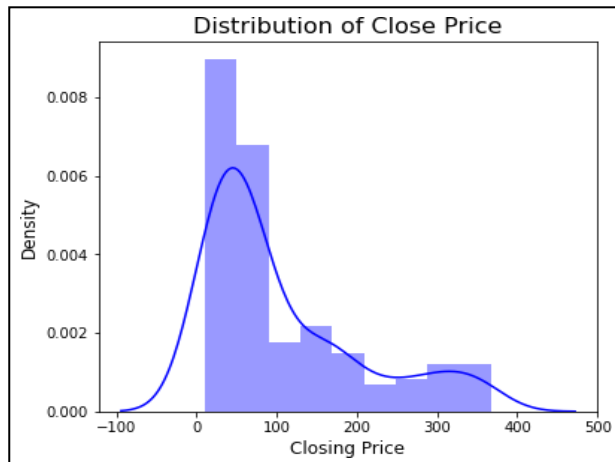
From above plot, there may strong correlation between each feature.

Distribution plot of dependent variable which is closing price of stock.

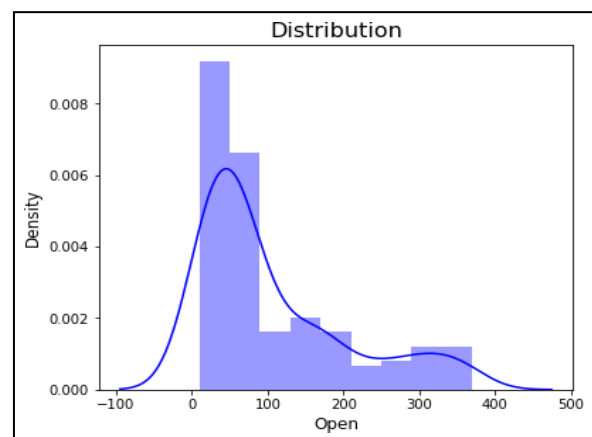
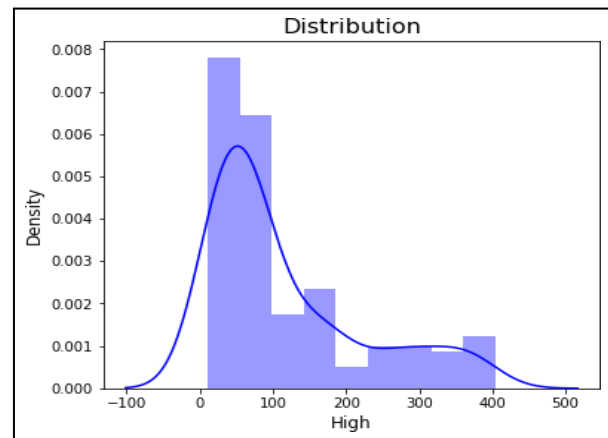


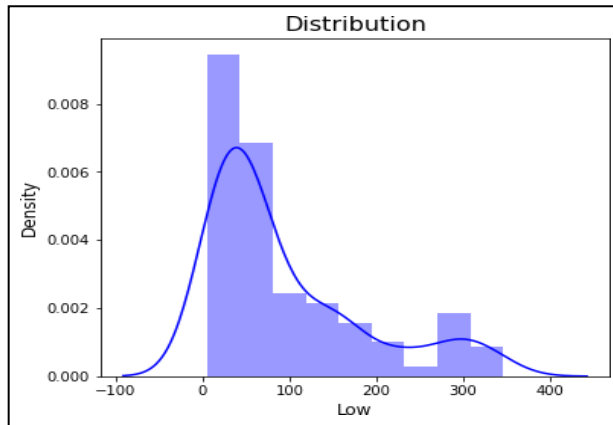
Now, distribution of closing price is normal distribution.

Distribution plot for each independent variable which is open, high and low price of stock.

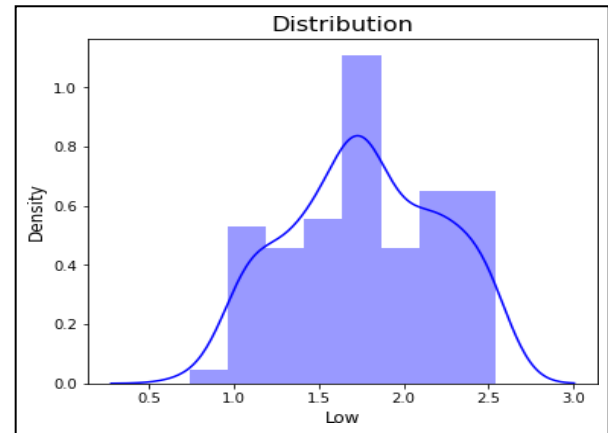


Distribution of closing price is right skewed. We need this distribution to be normal distribution for training algorithm. Log transformation applied over closing price to make it normal distribution.

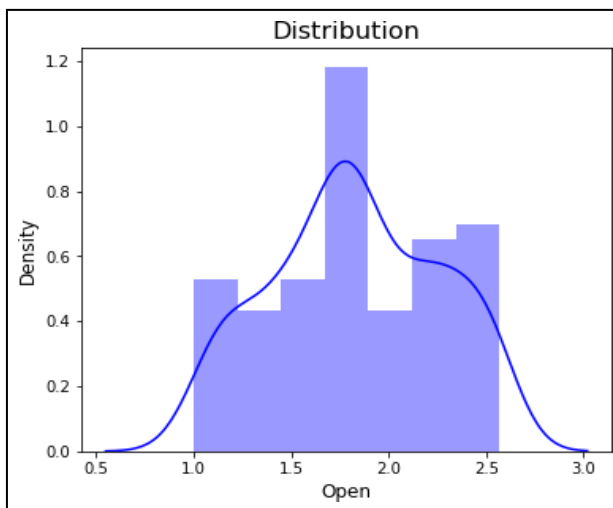




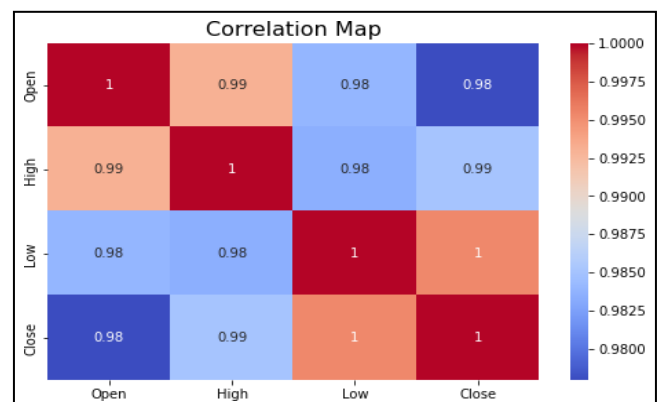
Distribution of opening price, high price and low price are also right skewed. Log transformation applied to make this distribution normal.



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



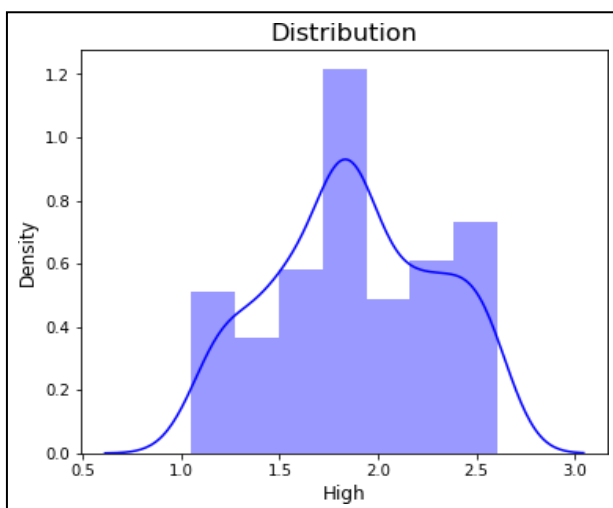
Correlation among the all features can be seen from heatmap.



From above map, all the features are strongly correlated with each other.

- **Standardization of features**

Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it.



The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

- **Splitting Data**

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

- **Fitting different models**

For modelling we tried various algorithms like:

- Linear Regression
- Lasso Regression
- Ridge Regression
- Elastic Net

- **Hyperparameter Tuning & Cross Validation for better accuracy**

Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting.

4. Algorithms:

4.1 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.

Following are the evaluation metrics after fitting data into linear regression model:

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

4.2 Lasso Regression:

Lasso (least absolute shrinkage and selection operator) 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.

Following are the evaluation metrics after fitting data into lasso regression model:

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

4.3 Ridge Regression:

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. 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.

Following are the evaluation metrics after fitting data into ridge regression model:

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

4.4 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.

Following are the evaluation metrics after fitting data into elastic net model:

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

5. 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.

5.1 CV & tuning on Lasso Regression:

The best fit alpha value is found out to be 0.01

The negative mean squared error for is -0.035

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

5.2 CV & tuning on Ridge Regression:

The best fit alpha value is found out to be 10.

The negative mean squared error for is -0.035

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

5.3 CV & tuning on Elastic Net:

The best fit alpha value is found out to be 0.01

The negative mean squared error for is -0.035

References-

1. StackOverFlow
2. GeeksforGeeks

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

6. 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

7. Conclusion:

1. Maximum accuracy of 82% achieved.
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.
4. Close, Open and high price of stock are strongly correlated with each other.