# Yes Bank Stock Closing Price Prediction

Janhavi Shembade, Data Science Trainee @almabetter

# ABSTRACT:

The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening a new business or the need for a high salary career.

Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions.In this project we attempt to implement a machine learning approach to predict stock prices.

We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades.

In this project we are using Machine learning, which give a prediction of various aspects of a particular stock or an index, such as future values of the opening price, closing price, index value etc.This will help investors and traders make better and faster decision

*Keywords- Stock market; Prediction; Machine learning; Artificial neural network*

**INTRODUCTION:**

Stock market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country.Both investors and industry are involved in stock market and wants to know whether some stock will rise or fall over a certain period of time.

The stock market is the primary source for any company to raise funds for business expansions. It is based on the concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for company's stock is low then the company share price decreases

The successful prediction of a stock's future price could yield significant profit. This is done using large historic market data to represent varying conditions and confirming that the time series patterns have statistically significant predictive power for high probability of profitable trades and high profitable returns for the competitive business investment.

.

**BUSINESS UNDERSTANDING:**

In this Finance word stock trading is one of the most important activity.It is an act trying to determine the future value of stock other financial instruments traded on a financial exchange.Basically quantitative trader with a lot of money from stock markets buy stock derivatives and equities at a cheap price and later on selling them at a high price..

**PROBLEM STATEMENT:**

We propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then use the acquired knowledge for an accurate prediction.

Basically quantitative traders with a lot of money from stock markets buy stock derivatives and equities at a cheap price and later on selling them at a high price.

**The main goal of the project is to:**

This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by most of the stockbrokers while making the stock predictions. The programming language used to predict the stock market using machine learning is Python.

**DATA DESCRIPTION:**

The data description phase starts with an initial data collection and proceeds with activities in order to get familiar with the data. Identifying data quality problems, discovering first insights into the data and detecting interesting subsets to form hypotheses from hidden information are activities of this step. Data which is collected from a bank to get analysed, involves usage details of stocks. The data was taken from Yes Bank which is a Financial bank domain. It had 185 rows and 5 columns. Most columns are related to stocks. Other column was indicative of service usage by the date. Based on the business understanding of the data 5 columns was chosen to analyse the data

**DATASET PREPARATION:**

The stock dataset from Yes Bank contains 5 features and 185 observations. The feature ‘Close’ shows stocks' Close price based on existing conditions. Below Table shows the data features.

**Data-set description**

|  |  |
| --- | --- |
| **Feature Name**  Date  Open  High  Low  Close | **Type**  object  Int64  Int64  Int64  Int64 |

**FEATURE BREAKDOWN:**

**Date**: The stock day date is mentioned in categorical type

**Open**:Stock Opening Price

**High:** Highest price of stocks

**Low:**  Lowest price of stock

**Close:** Closing price of stock

**EXPLORATORY DATA ANALYSIS:**

If we want to explain EDA in simple terms, it means trying to understand the given data much better, so that we can make some sense out of it. It was also used to produce a value distribution and identify missing values, and outliers.

EDA is a process of examining the available dataset to discover patterns, spot anomalies, test hypotheses, and check assumptions using statistical measures. In this chapter, we are going to discuss the steps involved in performing topnotch exploratory data analysis

In statistics, A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task.EDA in Python uses data visualization to draw meaningful patterns and insights

* **DATA ANALYSIS:**

This is one of the most crucial steps that deals with descriptive statistics and analysis of the data. The main tasks involve summarizing the data, finding the hidden correlation and relationships among the data, developing predictive models, evaluating the models, and calculating the accuracies. Some of the techniques used for data summarization are summary tables, graphs, descriptive statistics, inferential statistics, correlation statistics, searching, grouping, and mathematical models.

* **DATA CLEANING**

After completing the Data Sourcing, the next step in the process of EDA is Data Cleaning. It is very important to get rid of the irregularities and clean the data after sourcing it into our system.

Irregularities are of different types of data.

* Missing Values:-So In this dataset we don't have any missing values
* Incorrect Format:-The “Date” feature only in object data type overall having int data type
* Incorrect Headers:- All headers are in good format
* **DATA TRANSFORMATION:**

Data transformation is the process of normalizing and aggregating the data to further improve the efficiency and accuracy of data mining.

So here we are using mean average aggregation which takes the mean of all “OHL” and creates new\_features.

* **DATA DEDUPLICATION:**

It is very likely that your dataset contains duplicate rows. Removing them is essential to enhance the quality of the dataset.

So there is no duplicate value present

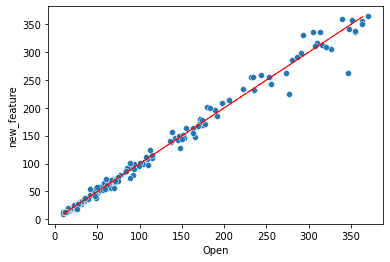
* **BIVARIATE ANALYSIS:**

If we analyze data by taking two variables/columns into consideration from a dataset, it is known as Bivariate Analysis.

* **a)Numeric-Numeric Analysis:**

Analyzing the open variables and Date variable from a dataset is known as numeric-numeric analysis. We can analyze it in three different ways.

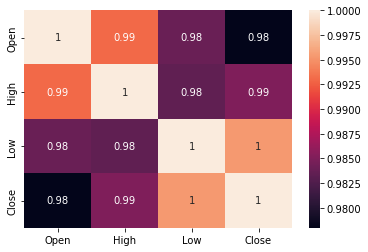
* Scatter Plot
* Pair Plot
* Correlation Matrix



this graph shows the relation between dates and open

* **CORRELATION AMONG VARIABLES**:

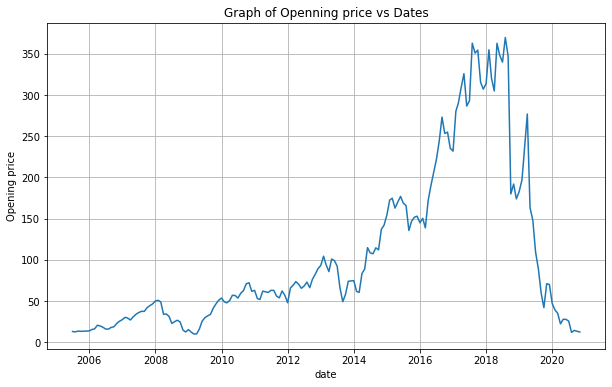
In words, the statistical technique that examines the relationship and explains whether, and how strongly, pairs of variables are related to one another is known as correlation.Each variable is highly correlated with each other.



So from graph its seem like there is multicollinearity is present.To reduce this we need to use regularization method.this is good when we have to predict each dependent variable

* **GRAPHICAL REPRESENTATION OF THE RESULTS:**

This is an essential step as the result analyzed from the dataset should be interpretable by the business stakeholders, which is one of the major goals of EDA.



So Opening price of a stock and closing price of a stock have the same results.Both have a max peak point in the year of 2018 and fall down year in 2020.

* **REGULARIZATION:**

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

The commonly used regularization techniques are :

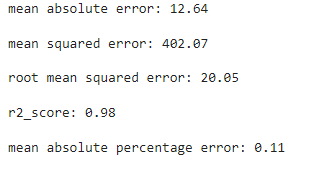
1. L1 regularization (Lasso Regression)
2. L2 regularization (Ridge Regression)

1.L1 Regularization:It also Known as **(Least Absolute Shrinkage and Selection Operator)**

**Lasso Regression** adds *“absolute value of magnitude”* of coefficient as penalty term to the loss function(L).

**Loss function with L1 regularization :**

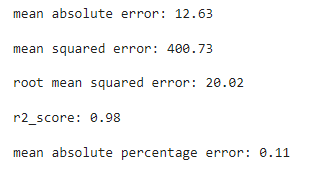
**L = y log (wx + b) + (1 - y)log(1 - (wx + b)) + lambda\*||w||1**



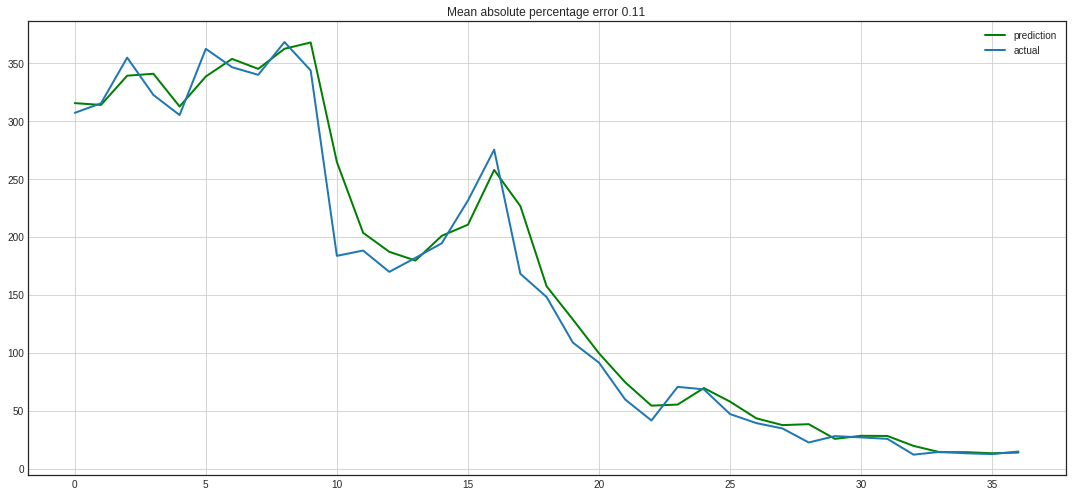
2.L2 Regularization :It also Known as **Ridge regression** adds “*squared magnitude*” of coefficient as penalty term to the loss function(L).

**Loss function with L2 regularization :**

**L = y log (wx + b) + (1 - y)log(1 - (wx + b)) + lambda\*||w||22**

****

**GRAPHICAL ANALYSIS FOR BOTH REGULARIZATION:**



* So from the above graph of yes bank open price and yes bank closing price it seems like both graphs have the same results.
* Opening price is started to increasing from 2016 and having its peak point in the year of 2018 and laying down is started from in range year of 2019-20

**CONCLUSIONS:**

From the Graph of Stock opening price and stock closing price have the same results. From the new\_feature attribute and close attributes seem like there is high correlation of each variable.

Bivariate analysis shows high correlation of close price with each other Graph of “closing price vs date” and “open price vs dates” shows that from the date range of 2018-19 there was falling down of stock and it becomes 0 in 2020.

We are implemented a linear regression model which gives the accuracy of 98% and visualize the linear regression model gives best fit model