Alphabet\_Inc\_Stock\_Market\_Prediction

Aim:

Stock market prediction of Alphabet Inc.

* In this project , I will analyse and perform predictive analysis on the historic data of Alphabet Inc.,and perform a classification or
* regression task to predict future price or return.
* Alphabet Inc. is an American multinational conglomerate headquartered in Mountain View , California.It was created through a corporate restructuring of Google , and became the parent company of Google and several former Google subsidiaries

Data Collection:

* Alpha vantage APIs is used for Data Collection.
* ‘data’ and ‘meta\_data’ are extracted from the obtained API.
* ‘Meta\_data’ store the meta information of the data like :

Information: Daily Prices(open,high,low,close

and volumes),

Symbol: ‘GOOGL’,

Last Refreshed: ‘xxxxx’

Output Size: ‘Full size’,

Time Zone:’US/Eastern’

Data Preprocessing:

* Having Total of 6 columns of which 5 are float type while 6th one is datetime type.
* There is no null values in the dataframe.
* Index is based on the date and so i change the index date into a feature and arrange all the data in ascending order according to the date.
* We can drop ‘Volume’ Feature from the dataframe as it is is the number of shares or contracts traded in a security or an entire market during a given period of time which gives an investor an idea of the price action of a security and whether he should buy or sell the security.

Visualization:

* From histogram plot we get an idea of distribution of data, ‘open’,’close’,’high’ and ‘low’ have the same type of data distribution.
* From pairplot all the four features vary linearly with each other as I have already told you that they have same type of data distribution.
* So I will use any one feature for the stock prediction as it will of no use to include all the features for prediction.

Models:

* Models used in the stock prediction are :

1. Moving Average:

* About

The predicted closing price for each day will be the average of previously observed values.

for each subsequent step, the predicted values are taken into consideration while removing the oldest observed value from the set.

* Prediction

On finding the RMSE value we get

RMSE=171.80

Which is quite high error but it is okay for with respect to model simplicity.

But from the trend plot of predicted value , the predicted value follow the same trend as the real value, first increases then decreases.

2.Linear\_Regression:

* About

We will first create a separate dataset, so that any feature engineering based on the model doesn't affect the original dataset.

We will first create a separate dataset, so that any feature engineering based on the model doesn't affect the original dataset.

* Feature Engineering

we can extract more features from 'date' column other than 'day' , 'month' and 'year' like :

'Dayofweek',

'Dayofyear',

'Is\_month\_end' or 'Is\_month\_start' etc.

The above features are created using (fastai structured module).

* Prediction

Linear Regression perform poorly than Moving Average technique as RMSE value for Linear Regression is higher than Moving Average.

Linear Regression is simple technique and easy to interpret.

But it is prone to overfitting it performs well for problem where the independent features are useful for determining the target value.

3.Auto\_Arima

* About

Popular Statistical method for time series forecasting.

It process past values to predict the future values.

* Prediction

The RMSE Value is 70.59 for

Auto-Arima Model.

An auto ARIMA model uses past data to understand the pattern in the time series ,using these values the model captured an increasing trend in the series.

These predictions are still not close to the real values but are far better than previously implemented machine learning model.

the model has captured a trend in the series, but does not focus on the seasonal part.

4.LSTM

* About

use for sequence prediction problem and is extremely effective.

it is able to store past information that is important , and forget that is not.

* Prediction

The RMSE value is very small which is far much better than previously implemented machine learning models.

From the trend plot one can easily visualize that predicted value follow exactly the same trend as the target variable.

Conclusion:

stock prices are a function of information and rational expectations, and that newly revealed information about a company's prospects is almost immediately reflected in the current stock price.

stock price is also affected by the factors like demonetization or merger/demerger of the companies. There are certain intangible factors as well which can often be impossible to predict beforehand.