**Oil Price Prediction**

**Dataset: -**

<https://in.investing.com/commodities/crude-oil-historical-data>

**Problem Statement**: -

Xyz company acquired the following data of oil prices changes occurred daily,

Make a model to predict the prices for months/dates to help the customers to get the price hike and decrease in prices to invest in particular stocks

**Steps to perform**

**Import data**: Used Octoparse for web-scrapping. Had to extract the data in two different csv files because at one time only limited data could be extracted, after importing we merged the files into one and then formed a single excel workbook.

**EDA:**

1.Removing Columns

2.Checked for Duplicate Values

3.Converting Time Parameter with Correct Data type

4.Performed Descriptive Stats to look for the initial insight of data

5.Treating outliers on business perspective

**Data visualization:**

1.Line Plot: To verify the Seasonality, Trend and Level

2.Histogram: Helps us to look for the distribution of the data

3.Box plot: To detect outliers

4. Lag plot: Quantifies the level of association with Yt and It’s lag versions.

**Data Transformation:**

To Remove noise and improve the signal of time series forecasting Depend on data set we selected best transformation like sqrt and log, Which helps us to make the distribution uniform.

**Split data into train and test:**

We consider old data for training purpose and few previous data for testing purpose, Later on we deploy model by joining training + validation to forecast the future.

**Forecasting methods:**

*Past pattern/trend is not similar to future we can apply following data driven Models*

1.Naive models

2.Moving Average smoothing

3.Exponential Smoothing

* Simple Exponential Smoothing
* Double Exponential Smoothing (Holt’s Method)
* Triple Exponential Smoothing (Holt’s Winter Method)

4.Auto Regression Models

* AR(p) Model
* MA(q) Model
* AR(p)MA(q) Model
* AR(p)I(d)MA(q) Model

**Model Evaluation:**

Evaluating Model Predictive Accuracy By using One of the Following metrics,

And Their value should be low as possible.

* Root mean squared error
* Mean absolute percentage error

**Hyperparameter Tuning:**

In Order to Increase the accuracy of model By using Grid Search CV we pick the best combination of p, d and q values in ARIMA Model**.**

**Deployment:**

After getting the generalized model we deploy the results for business with the help of ‘Streamlit’ and ‘Keras’.