Module of the project: Machine Learning to predict share prices in the Oil & Gas Industry

Purpose: preparing data dictionary

Abstract: The main objective of this module is to predict share price of different oil and gas companies based on past history of brent oil prices and share price history of different company like Shell (RDSB.L),BP (BP.L),Cairn Energy (CNE.L),Premier Oil (PMO.L),Statoil (STL.OL),ENGIE (ENGI.PA),Schlumberger (SLB.PA).

we have gathered few data sets of brent oil prices in daily manner and share prices of above mentioned companies.(i will attach those datasets as well with this document ,please go through them).

BRENT OIL PRICE DATASET:

We have two columns in this oil price dataset which contains date and price of the brent oil on daily basis from 2000 to 2020.By this project we would like observe the data of oil prices trends and patterns.by that observed analysis, we can draw some conclusion part of the data how it has changed and what factors will influence.by that deep analysis of the data,i believe we can predict oil prices with best accuaracy.

dataset refference:**Oil price dataset** from the U.S Energy Information administration.

My head of the oil price dataset:

	date	oil_price
2	1987-05-20	18.63
3	1987-05-21	18.45
4	1987-05-22	18.55
5	1987-05-25	18.6
6	1987-05-26	18.63

The stock market price dataset:

In this share price data set of few companies as mentioned above, we have different columns for each company share price like date, open,high,low,close,adjusted close and volume. Generally for any share price market we will have these kind of columns only which normally means the opening share value of that market ,low and high prices in that timeframe and closing value by the time of trade time off. volume column gives us how many shares has been sold in that time frame, based on that volume the company would adjust the share value and gives us adjusted close value. those are the columns we have in the share price data sets of few companies commonly.

The route map of this module to meet the objective :

Luckily we got our datasets which are in proper order, nothing much to do with data cleaning except few of them.

Brent dataset info:

Total two columns date and oil price.

Total 7641 rows in date and oil price columns with 0 null values.

Firstly we would like to combine all companies share price and brent oil prices into one master data frame. then we start plotting different plots between different companies to observe trends and patterns. The **pairplot shows all the pairwise relationships in a dataset and the univariate distribution** of the data for each variable. It gives us a reasonable idea about variable relationships .we can plot relationship between oil price and different companies share price by plotting methods in seaborn library to see how the share price evolves with respect to time and oil prices. Thorugh those method we will be able to identify how sensitive our data in stock prices of different companies w.r.t oil prices. Then i believe by those deeply obseved patterns we will be able to predict some prices of oil and share prices of the companies with best accuaracy.