**Crude Oil Price Prediction**

**Abstract**

Crude oil is the most important commodity in the world. Trillions of dollars of oil are being traded between counties every year and it is the underlying asset for thousands of stocks, futures, options and other financial vehicles. Machine learning has been proved to be an effective method to help predict oil price change, assisting traditional price forecasting approaches.

**Problem Statement:**

The goal of the project is to develop forecasting machine learning model to predict the oil prices based on the historical data. Oil is a product that goes completely in a different direction for a single market event as the oil prices are rarely based on real-time data, instead, it is driven by externalities making our attempt to forecast it even more challenging. As the economy will be highly affected by oil prices our model will help to understand the pattern in prices to help the customers and businesses to make smart decisions.

**Steps to perform**

1. Data Collection:

Collect the crude oil prices with dates from various websites by using web scrapping techniques and preprocess the data by preparing the raw data and making it suitable for a machine learning model. Like getting the dataset, importing libraries, importing datasets, finding missing data, null values, outliers, etc.,

1. Exploratory Data Analysis (EDA):

Analyze the data and discover trends, patterns, or check assumptions in data with the help of statistical summaries and graphical representations. Explore and visualize the data to investigate the crude oil prices.

1. Feature Engineering:

Feature engineering is the pre-processing step of machine learning, which extracts features from raw data. It contains mainly four processes:

1. Feature Creation
2. Transformations
3. Feature Extraction
4. Feature Selection
5. Splitting the Train and Test data:

Split the data to train and test for model training. To estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model.

1. Model Building:

Choose the suitable algorithm for the model and train it according to the requirement. Check the accuracy of the model through the test data and apply the model. Like Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), LSTM (Deep Learning), Exponential Smoothing (ES), etc.,

1. Model Evaluation:

It is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Like Accuracy, Precision, recall, MSE, RMSE, Confusion Matrix, AUC-ROC Curve, True positives, True Negatives, False Positives, False Negatives, F1 Score, etc., Hyper parameter tuning- If the model performance is not showing any good results, then consider hyper parameter tuning to tweak model performance for optimal results. This process is an essential part of machine learning, and choosing appropriate hyperparameter values is crucial for success.

1. Deployment:

Once everything is done, then deploy the model using Stream lit, Flask or RShiny.