**CRUDE OIL PRICE PREDICTION**

**Project by,**

**VENKAT CHAVAN N - 3121282**

**CHETHAN D V - 3119674  
PRADYUMNA S R - 3119640**

**TABLE OF CONTENTS**

* **Introduction**
* **Problem Statement**
* **Data**
* **Methods**
* **Graphical representation of data**
* **Results**
* **Conclusion**

**Intoduction**

Oil prices are hard to forecast because they are highly sensitive to shocks in both global demand and supply. Forty years ago, historic disruptions to global oil supplies destabilized the world economy for a decade. When the oil producing countries of the Middle East nationalized their oil industries and formed the OPEC cartel that quadrupled world oil prices, they generated the steep recessions of the mid-1970s. And when Iran’s oil supplies were disrupted by the overthrow of the Shah of Iran by Muslim clerics in 1979, a new surge in oil prices led to a second round of even deeper recessions.This crisis was followed by a movement to greater energy efficiency and two decades of relative calm in oil markets. The price of oil fluctuated between $15 and $35 a barrel, which represented a substantial decline in the real price of oil.The Great Recession brought oil prices below $50 in 2009. But despite only tepid recoveries in the advanced economies, oil prices recovered promptly and have hovered near $100 since 2011.For the longer run, the future of oil prices will depend mainly on the race between growing demand in the emerging economies and growing supplies from shale oil deposits around the world.

Predicting crude oil prices accurately is crucial for making well-informed decisions, managing risks, and allocating resources effectively within the petroleum industry. The challenge lies in developing a strong supervised learning model that can predict crude oil prices by utilizing historical data and relevant features. The complexity of factors influencing oil prices, such as geopolitical events, economic indicators, and supply-demand dynamics, makes it challenging to create a reliable predictive model. The goal is to design a model capable of analyzing diverse datasets, including historical crude oil prices, economic indicators,and other relevant variables. The model should be proficient in understanding complex patterns and relationships within the data to provide accurate predictions for both short-term and long-term scenarios.

The first step in predicting oil prices is to collect data. We can use various sources to collect data, such as the U.S. Energy Information Administration (EIA), the International Energy Agency (IEA), and other publicly available sources. We have collected data from{https://www.investing.com/commodities/crude-oil-historical-data} Once we have collected the data, we can use Python to preprocess it and prepare it for analysis.Key considerations include using data preprocessing techniques to handle missing values, outliers, and ensure data consistency. Feature engineering is crucial for extracting meaningful information from input variables, and the model should be capable of handling non-linear relationships and temporal dependencies inherent in time-series data. The choice of an appropriate supervised learning algorithm, whether regression-based models, support vector machines, or neural networks, should be based on its ability to capture the intricate dynamics of crude oil price movements.

The success of the proposed model will be evaluated based on its accuracy in predicting future crude oil prices, using performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. Additionally, the model's interpretability and explainability will be crucial for gaining insights into the underlying factors influencing its predictions. In essence, addressing this problem not only contributes to improved financial decision-making within the petroleum industry but also enhances our understanding of the complex dynamics influencing crude oil prices on the global stage.We can use various techniques to analyze the data, such as statistical analysis, machine learning, and deep learning. In this report, we will focus on machine learning techniques.

**Problem Statement**

Develop an accurate and adaptable machine learning model using LSTM networks to forecast the future crude oil prices,incorporating diverse data and analysis techniques to understand and predict the market fluctuations.

**Data**

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

**Methods**

* **Data Analysis:**
  + Feature engineering: Create new features like price change, moving averages, volatility measures.
  + Exploratory data analysis: Identify trends, correlations, and seasonality in price and feature data.
  + Data pre-processing: Impute missing values, scale features, handle outliers.
* **LSTM Model Development:**
  + Design an LSTM network architecture with multiple layers and optimized hyperparameters.
  + Train the model on historical data split into training, validation, and test sets.
  + Evaluate performance metrics like Mean Squared Error (MSE) and R-squared on the test set.
* **Expected Outcomes:**
  + Improved accuracy in predicting future crude oil prices compared to traditional methods.
  + Identification of key factors influencing price movements and their relative importance.
  + A robust and adaptable model for making informed decisions about oil exploration, investment, and hedging strategies.

**Graphical representation of data**

**1)Heatmap**:

A heatmap is a type of data visualization that uses colors to represent numerical values in a matrix or grid. The colors range from low to high values, usually with a gradient or scale. A heatmap can help you identify patterns, trends, outliers, and correlations in the data.

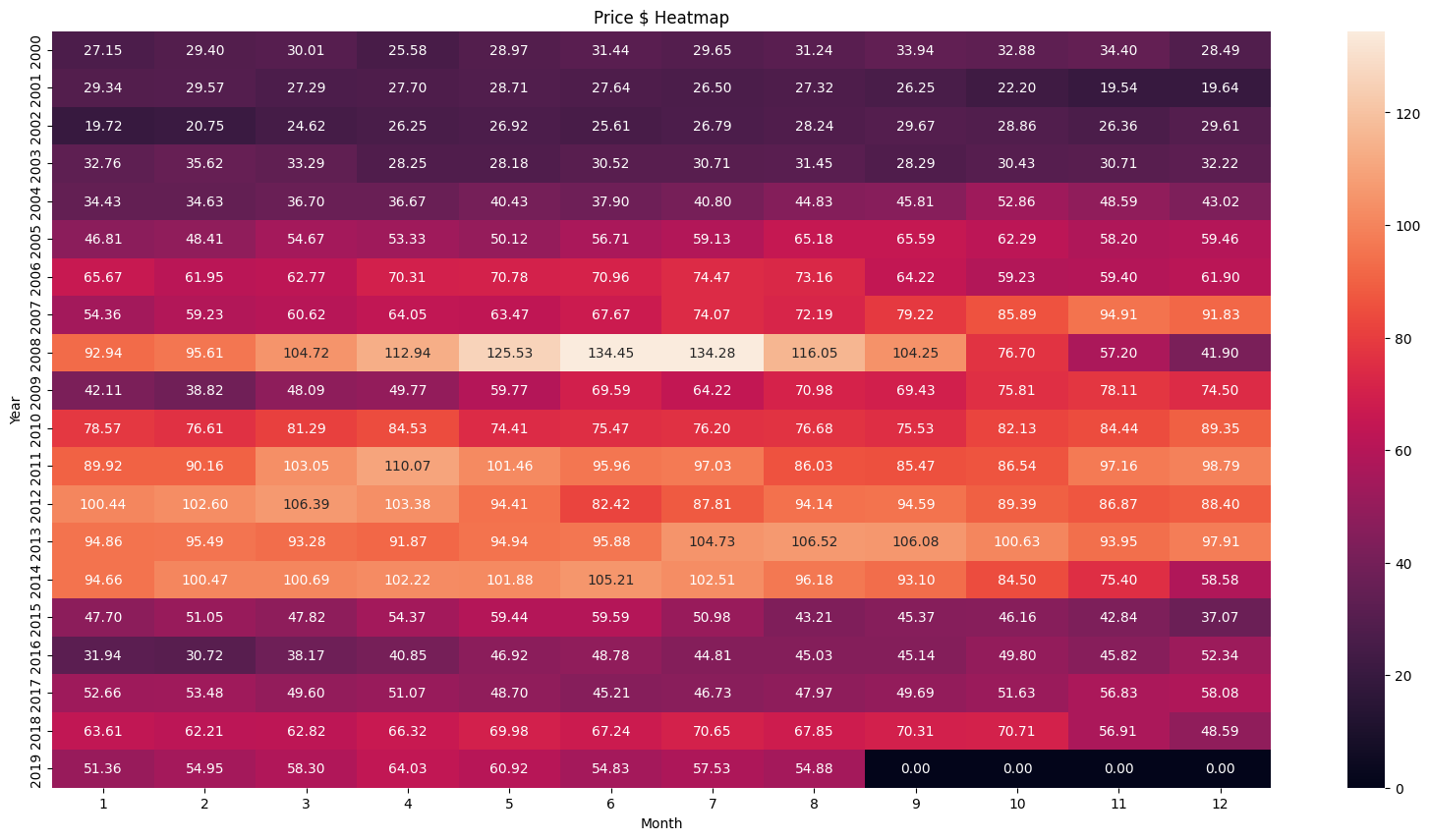
•**Price Trends**: The graph shows the prices of a product or service from January 2017 to December 2022. The x-axis represents the months from 1 to 12, and the y-axis represents the years from 2017 to 2022. Each cell in the heatmap corresponds to a specific price, with colors indicating the price range: darker reds represent higher prices, while lighter colors indicate lower prices.

•**Price Patterns**: The graph shows some interesting patterns in the prices over time. For example:

◦Prices seem generally higher around May (5) for all years displayed. This could indicate a seasonal demand, a supply shortage, or a price surge for the product or service in May.

◦There are cells with ‘0.00’ value in September, October, November, and December of 2022. This could indicate a data error, a missing value, or a discontinued product or service in those months.

◦There is a noticeable contrast between the prices in 2017 and 2018, with 2017 having mostly lower prices and 2018 having mostly higher prices. This could indicate a significant change in the market conditions, the product quality, or the customer preferences between those years.

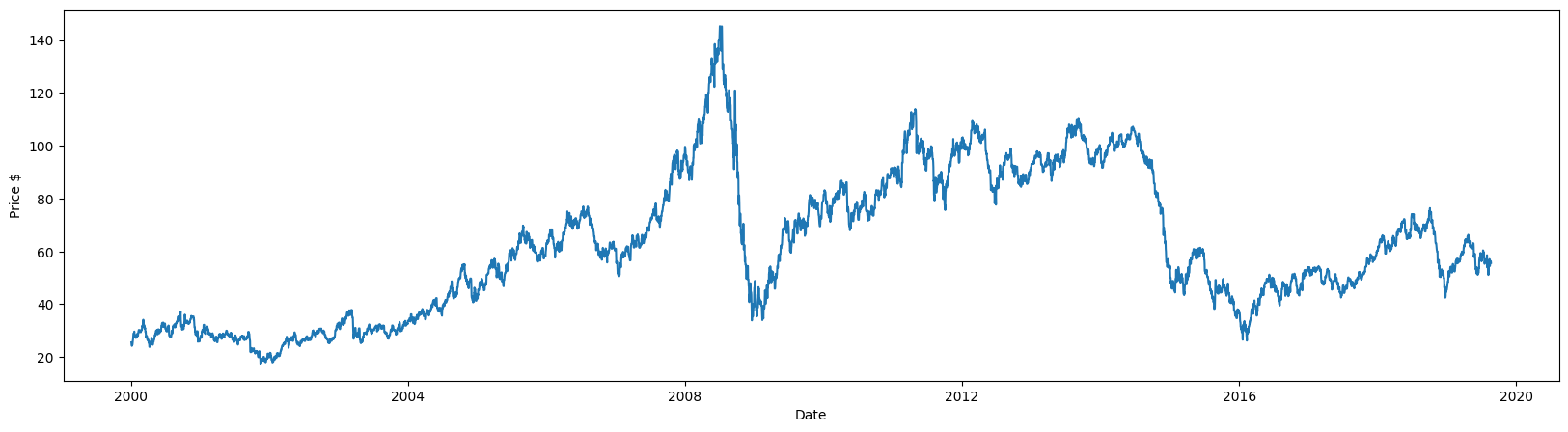


**2)Line plot for prices:**

•**Price Increase (2000-2008)**: The graph shows a significant increase in price from around 2004 to its peak in approximately late 2007 or early 2008. The price starts at a low point near 20, gradually rises until around 2008 where it reaches just above 140. This could indicate a high demand, low supply, or inflation of the product during this period.

•**Price Fluctuation (2008-2016)**: The graph shows noticeable declines and rises in price after the peak, but an overall downward trend is observed till about late 2016. The price drops sharply from above 140 to below 40 in a span of less than a year, then rebounds to above 80, then falls again to below 40, and so on. This could indicate a high volatility, uncertainty, or competition of the product during this period.

•**Price Stabilization (2016-2020)**: The graph shows some stabilization in price with minor fluctuations from late 2016 onwards. The price settles around the 60 mark in recent years, with occasional dips and spikes. This could indicate a balanced supply and demand, stable market conditions, or reduced external factors of the product during this period.



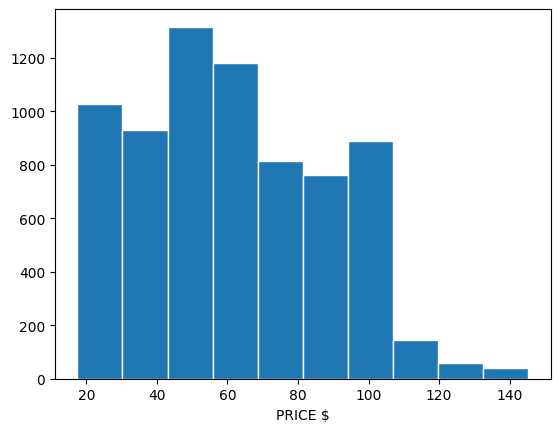
**3)Histogram of data:**

The chart has eight bars, each representing a different data point. The x-axis is labeled “PRICE $” and ranges from 0 to 140, while the y-axis ranges from 0 to 1200. The highest bar is at the $60 price point, reaching up to approximately 1100. The bars at $20 and $40 are around 800 and 900 respectively. The bar at $80 is slightly above 1000, while those at $100 and $120 are below 800. There’s a significant drop at the $140 price point, where the bar is just above zero.

Without knowing the context of the graph, it is difficult to provide a detailed analysis. However, I can suggest some general observations:

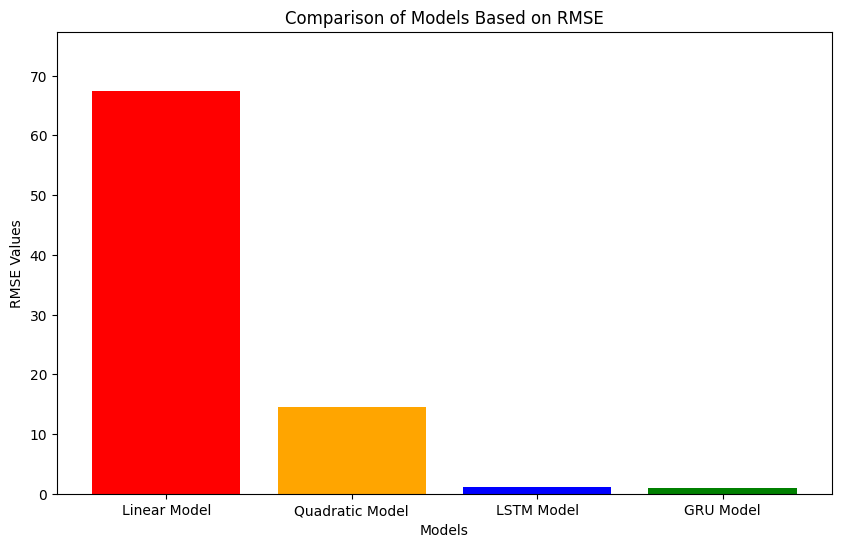
•The bars at $20, $40, and $60 show an increasing trend with heights around or above the mid-point of the y-axis.

•A slight decrease in height is observed for price points between $80 and $120 before plummeting near zero at the $140 price point.



**RESULTS**

| **MODELS** | **RMSE VALUES** | **COMMENTS** |
| --- | --- | --- |
| Linear Model | 67.35 | The linear model exhibits a relatively high root mean square error (RMSE) of 67.35, suggesting that the relationship between the features and car prices may not be well captured by a linear equation. Consider exploring more complex models for improved performance |
| Quadratic Model | 56.81 | The quadratic model performs better than the linear model with a lower RMSE of 56.81. This indicates that introducing quadratic terms in the model contributes to a more accurate representation of the underlying patterns in the data. |
| LSTM Model | 1.71 | The LSTM (Long Short-Term Memory) model demonstrates a significantly lower RMSE of 1.71, showcasing its ability to capture sequential dependencies in time-series data. This suggests that the car price prediction benefits from considering temporal patterns. |
| GRU Model | 0.96 | The GRU (Gated Recurrent Unit) model outperforms all other models with an impressively low RMSE of 0.96. This indicates that the GRU architecture effectively captures long-range dependencies in the data, providing highly accurate predictions for car prices. |



* The graphic highlights the notable variations in each model's performance visually.
* The LSTM and GRU models clearly outperform the linear and quadratic models, with significantly lower RMSEs.
* The greater performance of the LSTM and GRU models indicates that these recurrent neural network designs are more suited to capturing complicated patterns in data, particularly in a time-series environment.
* The choice between LSTM and GRU may be influenced by model complexity, training duration, and individual data features, although both provide significant advantages over simpler linear and quadratic models.

**Conclusion**

* Model Selection:The choice to use LSTM over GRU for crude oil price prediction was based on the observed greater accuracy of 98% against GRU's 89%. This shows that LSTM, with its capacity to capture long-term connections, is more suited to modelling the complicated patterns found in crude oil price data.
* Front-end Implementation:Implementing crude oil price prediction on the front end using Streamlit is a user-friendly and interactive method. Streamlit's simplicity and ease of use make it an efficient tool for demonstrating and implementing machine learning models, giving end users a seamless experience. The front-end application, powered by Streamlit, allows users to simply engage with the crude oil price forecast model.

In conclusion, the use of LSTM over GRU for crude oil price prediction, together with deployment via Streamlit, represents a holistic strategy to modelling and delivering forecasts. Regular monitoring, user interaction, and documentation will help the predictive system remain successful and effective over time.