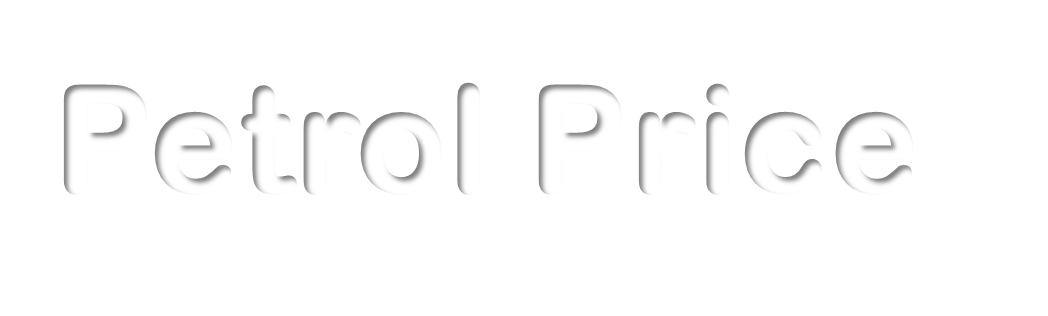
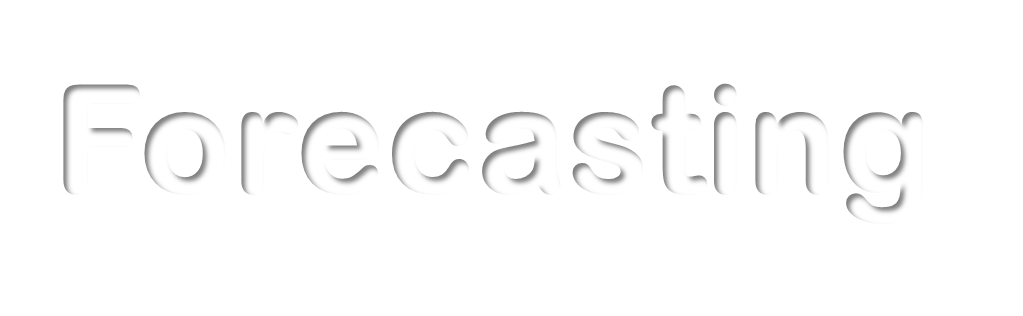


**Petrol Price**



**Forecasting**



Yudhisthir

|  |
| --- |
| **Objective:**  Development of a forecasting model to find forecasted price of petrol.  **Benefits:**  This model can be used to determine the petrol price in coming days based on present or historic datas. |

**Architecture**

**Data**

**Preparation**

**Model development**

Data

Preprocessing

Exploratory Data

Analysis

Feature

Engineering

Model

implementation

Model Training

Model

Evaluation

**Data validation and transformation**

|  |  |  |  |
| --- | --- | --- | --- |
| **DATA TYPE** | **NULL VALUES** | **NUMERICAL COLUMNS** | **CATEGORICAL COLUMNS** |
| Data type of columns is given in the schema file. It is validated when we insert the files into Database. | If any of the columns in a file have all the values as NULL or missing, we can fill it by some methods. | All the numerical features were standardized using Standard Scaler, preventing any data leakage. | Ordinal Encoding was used to treat categorical columns for the model in understandable way. |
| If data type is wrong, we can convert it using pandas library. | We can fill them by using mode of categorical columns or mean of numerical columns. | This process is done in pipeline for numerical features for the convenience of deployment. | This process is done in pipeline for categorical features for the convenience of deployment. |

**Data Ingestion**

⮚ Excel File- The dataset was imported from Excel File into python.

⮚ Data Frame was created using pandas.

**Model Training**

1.The data in database is imported to Jupyter notebook by using pandas.

2. In data preprocessing step, data is checked if there missing data, duplicate values.

**Model Selection:**

SARIMA and ARIMA model were used to predict the future value.

**Prediction**

The model is made in such a way to maximize the accuracy and also other performance metrics so that the predictions are as accurate as possible.

# Q&A

**What is the source of data?**

The exxcel file is source of data.

**What was the type of data?**

The data is numerical and date form type.

|  |
| --- |
| **What is the complete flow you followed in this project?**  Refer to 3rd slide for the process flow. |

|  |
| --- |
| What techniques were you using for data pre-processing?   * Visualizing relation of independent variables with each other and output variables * Checking for null values. * Checking for duplicate values. * Cleaning data and imputing if null values are present. * Scaling the data |

|  |
| --- |
| How training was done or what models were used?   * First, we started with data cleaning, EDA and feature engineering. Data type of columns were corrected by using pandas attributes. * Then Checked with ARIMA and SARIMA models. |