

# **APPLIED DATA SCIENCE**

## **(GROUP 2)**

**PROJECT TITLE-** Electricity Prices  
Prediction

PHASE-1(PROJECT)

**Data Source:** Start by obtaining a comprehensive dataset containing historical electricity prices. Ensure that the dataset also includes relevant factors like date, demand, supply, weather conditions, and economic indicators. The quality and completeness of your data will significantly impact the model's performance.

**Data Preprocessing:** Perform data cleaning to remove any outliers, errors, or inconsistencies in the dataset. Handle missing values appropriately, which may involve imputation or removal of incomplete data points. Convert categorical features into numerical representations using techniques like one-hot encoding or label encoding. Normalize or scale numerical features if necessary to ensure they are on the same scale.

**Feature Engineering:** Create additional features that can capture important patterns or trends in the data. For electricity price forecasting, time-based features like day of the week, month, and season can be valuable. Generate lagged variables to capture the effect of past prices on future prices. Lag features can help the model learn autocorrelations in the data.

**Model Selection:** Choose suitable time series forecasting algorithms. You've mentioned ARIMA and LSTM, which are both good options. Consider other models like Prophet, Exponential Smoothing methods, or even hybrid approaches that combine multiple models.

**Model Training:** Split your dataset into training, validation, and test sets. The training set is used to train the model, the validation set helps tune hyperparameters, and the test set evaluates the final model's performance. Train the selected model using the preprocessed data. Experiment with different hyperparameters to find the best configuration.

**Evaluation:** Assess the model's performance using appropriate time series forecasting metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), or others. Compare the model's predictions against the actual electricity prices on the test dataset. Visualize the model's predictions and actual values over time to gain insights into its performance and any potential biases or errors.

**Iterate and Refine:** Based on the evaluation results, iterate on the model design, feature engineering, or data preprocessing steps to improve performance. Consider using cross-validation techniques to ensure the model's robustness and generalizability.

**Deployment and Use:** Once you have a well-performing model, deploy it as a tool for energy providers and consumers to make informed decisions about consumption and investment. Continuously monitor the model's performance in a production environment and update it as needed to maintain accuracy. Throughout the process, it's essential to involve stakeholders, gather feedback, and keep the end-users' needs in mind to ensure that the predictive model effectively serves its purpose in assisting energy providers and consumers in decision-making