Electricity Prices Prediction

Team members:

Sahaana.k

Priyadharshini.s

Shobika.k

Rithika.s

PROBLEM STATEMENT

- Create a predictive model that utilizes historical electricity prices and relevant factors to forecast future electricity prices, assisting energy providers and consumers in making informed decisions regarding consumption and investment.

Selecting an appropriate data source is a critical step in your electricity price prediction project:

- **Data Availability:** Ensure that the dataset you choose is readily available and accessible for your project. It should include historical electricity price data as well as relevant factors like date, demand, supply, weather conditions, and economic indicators.
- **Scope of Data:** Ensure that the dataset covers a sufficiently long historical period. Having data spanning several years can be valuable for building accurate predictive models.
- **Data Format:** Make sure the dataset is in a format that is suitable for analysis. Common formats include CSV, Excel, or structured databases. Ensure that the data can be easily imported into your chosen data analysis tools or programming language
- **Data Updates:** Consider whether the dataset is regularly updated. Timely updates are essential if you plan to use the model for ongoing electricity price forecasting.

DATA PREPROCESSING:

- **Data Cleaning:**
- Handle Missing Values: Identify and deal with missing data points in your dataset. Common strategies include data imputation (filling in missing values with estimated values) or removing rows or columns with too many missing values.
- Remove Duplicates: Check for and remove any duplicate rows in the dataset to avoid redundancy.
- **Data Transformation:**
- Encoding Categorical Variables: If your dataset includes categorical variables (e.g., location, season), encode them into numerical values using techniques like one-hot encoding or label encoding.
- Scaling and Normalization: Scale numerical features to a common range, such as between 0 and 1, to ensure that they have equal weight during modeling. This is particularly important if you are using algorithms sensitive to feature scales, such as gradient-based methods.
- Log Transformation: If your target variable (electricity prices) or predictor variables have a skewed distribution, consider applying log transformations to make the data more normally distributed.

FEATURE ENGINEERING:

- Feature engineering involves creating new, meaningful features from existing data to improve a predictive model's accuracy. Techniques include extracting date components, generating lag features, aggregating statistics, including weather and economic indicators, and creating interaction terms. It's an iterative process guided by domain knowledge and data analysis, aiming to enhance model performance while avoiding overfitting. Regularly validate feature importance to select the most informative variables.

Model selection:

- If your electricity price data exhibits clear, linear trends and seasonality, and you prioritize interpretability, consider starting with ARIMA
- Choose ARIMA for straightforward, linear time series data with clear trends and seasonality. Opt for LSTM when dealing with complex, nonlinear data, including multiple variables and irregular patterns. Consider combining both approaches for enhanced forecasting performance. Experiment with various methods to find the best fit for your electricity price prediction project.
- LSTM, a type of recurrent neural network (RNN), is suitable for more complex time series data, including multivariate time series and sequences with long-range dependencies.

Model training:

Use the preprocessed dataset, which includes features and the target variable (historical electricity prices), to train the selected model (e.g., ARIMA or LSTM).

- Depending on the algorithm, configure the model with appropriate hyperparameters and settings.
- Fit the model to the training data, allowing it to learn the underlying patterns and relationships in the time series data.

Evaluation:

- After training, use the trained model to make predictions on a separate validation or test dataset. This dataset should be unseen during the training phase.
- Calculate relevant time series forecasting metrics to assess the model's accuracy and performance. Common metrics include:
  - **Mean Absolute Error (MAE),Root Mean Squared Error (RMSE),Mean Absolute Percentage Error (MAPE),R-squared (R2)**
  - Interpret the results of these metrics to understand the model's strengths and weaknesses in forecasting electricity prices.
- Use the evaluation metrics to compare different models or iterations of the same model, and select the best-performing one for deployment or further refinement.