**ELECTRICITY PRICE PREDICTION**

**PHASE 1 – DOCUMENT SUBMISSION**

***Team members****:*

*Divya .K (411621104008)*

*Annie mercy.S (411621104003)*

*Barkavi.E (411621104006)*

*Jeevaraji.S (411621104017)*

*Lekhashree.S (4116211028)*

Project Title: Electricity Prices Prediction

****

**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.

**Problem Definition:**

 The problem is to develop a predictive model that uses historical electricity prices and relevant factors to forecast future electricity prices. The objective is to create a tool that assists both energy providers and consumers in making informed decisions regarding consumption and investment by predicting future electricity prices. This project involves data preprocessing, feature engineering, model selection, training and evaluation.

**Design thinking:**

1**.Data source:**

In electricity price prediction, the choice of data sources is critical for building accurate and robust predictive models. Here are some common data sources used in electricity price prediction

* **Historical Electricity Price Data:**

Historical price data is the backbone of any electricity price prediction model. It includes information about electricity prices at different time intervals (e.g., hourly, daily) for a specific geographical region or market.

* **Weather Data:**

Weather conditions have a significant impact on electricity demand and supply. Data sources such as temperature, humidity, wind speed, and precipitation can be used to account for weather-related factors that influence electricity prices.

* **Demand Data**:

Information on electricity demand patterns is crucial. This data can come from utility companies and may include historical demand levels at different times of the day, week, or year.

* **Supply Data:**

Data on electricity generation and supply sources are important. This can include information on power plant capacity, availability, and generation methods (e.g., renewable, fossil fuels).

* **Economic Indicators:**

Macroeconomic data, such as GDP growth rates, inflation rates, and industrial production figures, can influence electricity demand and, subsequently, prices.

**2.Data preprocessing:**

Data preprocessing is a crucial step in electricity price prediction as it ensures that the data used for modeling is clean, relevant, and suitable for the predictive task. Here are key data preprocessing steps in electricity price prediction.

* **Data Cleaning**:
* Handle missing values: Identify and fill in or impute missing data points in the price, weather, or other relevant datasets. Common methods include mean imputation or interpolation.
* Outlier detection: Identify and address outliers in the data, which can significantly impact model performance. Outliers in price data, for example, could distort predictions.
* **Time Series Transformation**:
* Ensure stationary data: Many time series models assume stationarity. You may need to perform differencing or other transformations to make the time series data stationary.
* Seasonal decomposition: Decompose the time series into its seasonal, trend, and residual components using methods like seasonal decomposition of time series (STL).
* **Handling Imbalanced Data:**

Electricity price data can sometimes be imbalanced, with periods of stable prices and occasional spikes. Consider techniques like oversampling or under sampling to balance the dataset.

* **Data Splitting:**

Split the data into training, validation, and test sets. Ensure that the time order is preserved to simulate real-world forecasting scenarios.

**3.Feature engineering:**

Feature engineering is a crucial step in electricity price prediction as it involves transforming raw data into informative features that can improve the accuracy of predictive models. Here are some common feature engineering techniques and considerations specific to electricity price prediction.

* **Time-Based Features:**
* Time of day: Extracting features like hour of the day, day of the week, or month can capture daily, weekly, and seasonal patterns in electricity prices.
* Holidays and special events: Including binary flags for holidays or significant events that affect electricity demand and pricing.
* **Lagged Variables:**
* Lagged electricity prices: Adding lagged (previous time steps) prices as features can help capture temporal dependencies and autocorrelation in price data.
* Weather data: Incorporating lagged weather variables such as temperature, humidity, and wind speed can account for weather-related effects on electricity demand and supply.
* **Price Spread and Volatility**:
* Price spread: Calculating the difference between high and low prices within a time period can indicate market volatility.
* Volatility measures: Incorporating metrics like historical price volatility or implied volatility can provide insights into price uncertainty.

**4.Model selection:**

Selecting the right model for electricity price prediction is crucial for accurate and reliable forecasts. Model selection involves choosing an appropriate machine learning or statistical model that can capture the underlying patterns in the data. Here are some commonly used models for electricity price prediction.

* **Time Series Models:**
* ARIMA (Auto Regressive Integrated Moving Average): ARIMA models are widely used for time series forecasting, including electricity prices. They can capture seasonality and trends in price data.
* Seasonal Decomposition of Time Series (STL): STL decomposes time series data into seasonal, trend, and residual components, making it easier to model each component separately.
* Prophet: Developed by Facebook, Prophet is designed for forecasting with daily observations that display patterns on different time scales. It can handle holidays and special events, which are relevant in electricity price prediction.
* **Machine Learning Models:**
* Random Forest: Random Forest is an ensemble learning method that can capture complex relationships in the data. It is often used when there are nonlinear patterns in electricity price data.
* Gradient Boosting (e.g., XGBoost, LightGBM): Gradient boosting algorithms are powerful for regression tasks. They can handle large datasets and capture both linear and nonlinear patterns.
* **Support Vector Machines (SVM):**
* SVMs can be used for regression tasks, but they are less common in time series forecasting for electricity prices compared to other models.

* **Proximity-Based Models**:
* k-Nearest Neighbors (k-NN): k-NN can be used for regression by averaging the k-nearest neighbors' values. It's simple but may not capture complex temporal patterns well.

**5.Model training:**

Training a model for electricity price prediction typically involves the following steps:

* **Data Collection:**

Gather historical data on electricity prices. This data should include timestamps, price values, and potentially relevant features like weather conditions, demand, or generation sources.

* **Data Preprocessing:**

Clean and preprocess the data. This involves handling missing values, normalizing or scaling the data, and encoding categorical variables.

* **Deployment:** Once satisfied with the model's performance, deploy it to make real-time predictions on electricity prices.
* **Monitoring and Maintenance:** Continuously monitor the model's performance in a production environment and retrain it periodically with new data to ensure its accuracy remains high.

**6.Evaluation:**

In electricity price prediction, evaluating the performance of your predictive model is crucial to ensure its accuracy and reliability. Here are some common evaluation metrics and techniques used for assessing the quality of electricity price predictions:

* **Mean Absolute Error (MAE):**

MAE measures the average absolute difference between the predicted and actual prices. It provides a simple and interpretable measure of prediction error.

* **Mean Squared Error (MSE):**

MSE calculates the average squared difference between predicted and actual prices. It penalizes larger errors more heavily than MAE, making it sensitive to outliers.

* **Root Mean Squared Error (RMSE):**

RMSE is the square root of MSE and provides a measure of the typical prediction error. It's in the same unit as the target variable (electricity price) and is commonly used in regression tasks.

**CONCLUSION:**

In Phase1, we have established a clear understanding of our goal: To predict the electricity price. We outlined a structured approach that includes Data source selection ,Data preprocessing ,Feature selection, Model selection, Model training and Evaluation. This sets the stage for our project’s successful execution in subsequent phases.