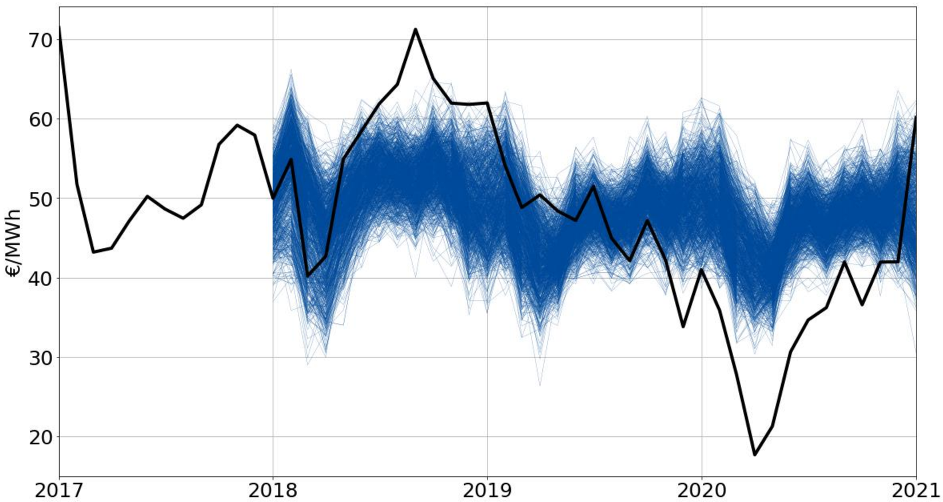
Electricity Prices Prediction

Team member:N.Sanjay
Register number:211521243135

Phase 1 :*Problem Definition and Design Thinking*

Problem statement:



To develop a model that can predict electricity prices for a specified future period based on historical data and relevant features, in order to enable stakeholders like utility companies, regulators, and consumers to make informed decisions.

Problem definition:

- Leverage data science techniques to forecast electricity prices based on historical and contextual data, aiming to improve the decision-making processes for utility providers, consumers, and policymakers.
- Predict the average wholesale prices of electricity for the next quarter/year, assisting utility companies in their long-term procurement strategies and contract negotiations.
- Model the effect of increased renewable energy integration on electricity prices over the next five years.
- Understand how demand-side responses to price changes, particularly during peak periods, and model the effect of demand response initiatives on future prices.

Project goals:

1. Accuracy Enhancement:

Achieve a high degree of prediction accuracy to reduce uncertainties in the electricity market.

2. Real-time Prediction Capabilities:

- **Objective:** Develop a model capable of producing real-time or near-real-time predictions for immediate operational decisions.

3. Long-term Forecasting:

- **Objective:** Provide reliable long-term forecasts that assist in planning and strategic decision-making for the upcoming months or years.

4. Adaptive Learning:

- **Objective:** Ensure the model can adapt to new data, recognizing and adjusting to emerging trends and patterns.

5. Feature Significance Analysis:

- **Objective:** Understand the primary drivers influencing electricity prices to inform policy decisions and market strategies.

6. Environmental Impact:

- **Objective:** Use price predictions to maximize the integration of renewable energy sources during periods of low costs, thereby reducing carbon emissions.

Design Thinking:

Applying design thinking to the electricity price prediction problem means focusing on a human-centered approach to address the challenges and needs associated with forecasting electricity prices.

Data Source:

1. Historical Electricity Prices:

- Hourly, daily, or monthly past electricity prices from power exchanges or utility companies.

2. Demand and Supply Data:

- Historical and real-time electricity consumption data.
- Installed capacity and the actual generation data from power plants.
- Reserve margins or backup capacity available.

3. Renewable Energy Data:

- Energy production from renewable sources like wind, solar, and hydro.
- The capacity factor of renewable installations.

4. Market Data:

- Fuel prices (natural gas, coal, oil), especially if these are primary sources for electricity generation in the region.
- Carbon or emission prices if there's a carbon trading system in place.
- Prices of Renewable Energy Certificates or similar instruments.

5. Economic Indicators:

- GDP growth rates, indicating overall economic activity and potential energy demand.
- Industrial production and consumption data, which could be a significant subset of total electricity demand.

Dataset Link:

https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction

Data preprocessing:

Data preprocessing is a crucial step in any predictive modeling project. For electricity price prediction, the data collected might be vast and varied, and it's essential to ensure that this data is clean, consistent, and ready for modeling. Here are the steps you should consider for preprocessing:

1. Data Cleaning:

2. **Handling Missing Values:**

Imputation: Replace missing values using methods such as mean, median, mode, or forward-fill/back-fill for time series data.
Deletion: Remove rows or columns with excessive missing values, especially if they don't contain crucial information.

2. Data Transformation:

Feature Scaling: Standardize or normalize features, especially if you're using algorithms sensitive to feature scales, like SVM or KNN.
Standardization (Z-score normalization)
Min-Max Scaling
Time Series Decomposition: For time series data, decompose it into trend, seasonality, and residuals.

3. Data Integration:

-

Merge Different Data Sources: Combine datasets such as historical prices, weather data, and demand data using common keys or timestamps.

-
-

Resolve Inconsistencies: Ensure that data from different sources matches in units, scales, and definitions. For example, ensure that demand is consistently measured in MW or MWh across datasets.

3. Data Encoding:

Categorical Encoding: Convert categorical variables into a format that can be provided to machine learning algorithms to improve accuracy and computational efficiency.

- **One-hot Encoding:** For nominal categorical variables.
- **Ordinal Encoding:** For ordinal categorical variables.
-

5. Data Visualization:

- Continually visualize data throughout the preprocessing steps using histograms, scatter plots, time series plots, and more to ensure that transformations are having the intended effect.

3.Feature Engineering:

Feature engineering is a pivotal step in the predictive modeling process. It involves creating new features from the existing ones, capturing additional information or patterns that can enhance the model's predictive performance. In the context of electricity price prediction, here are some feature engineering techniques and ideas:

1. Temporal Features:
Given the time series nature of electricity pricing data, time components can be very informative.

- **Hour, Day, Month, Year:** Extract these components from the timestamp.
- **Weekend vs. Weekday:** Prices might differ between weekdays and weekends.
- **Season:** Spring, Summer, Fall, Winter - Energy consumption patterns can vary by season.

2. Lag Features:
Past values and statistics can be indicative of future prices, especially in time series forecasting.

- **Lagged Prices:** Prices from previous hours, days, or weeks.
- **Rolling Means:** Average price over a specified window.
- **Rolling Standard Deviation:** Variability of prices over a specified window.

3. Demand & Supply Interaction:
The interplay between demand and supply can provide insights.

- **Demand/Supply Ratio:** A ratio of demand to supply at any given time.
- **Difference between Demand and Supply:** Absolute difference between demand and supply.

4. Historical Patterns:
Recurring patterns can emerge when looking at historical data.

- **Price Momentum:** The difference between the current price and the average price over the past 'n' days.
- **Historical Volatility:** Measure of price variability over a specified time period.

5. Market Features:
Influence from broader energy markets can impact electricity prices.

- **Fuel Prices:** Current prices of natural gas, coal, or oil if these are primary generation sources.
- **Carbon Credit Prices:** If there's a carbon trading system in place.

6. Infrastructure Features:
The status of power infrastructure can influence prices.

- **Maintenance Flags:** Indicators of major power plants or transmission lines undergoing maintenance.

Model selection:

Model selection for electricity price prediction should take into account the nature of the data (time series), the complexity of the problem, and the desired forecast horizon (short-term vs. long-term). Here are several models that can be considered, along with their potential advantages and use cases:

1. Classical Time Series Models:

ARIMA (AutoRegressive Integrated Moving Average):

-
    - Suitable for univariate time series data without external predictors.
    - Assumes linear relationships and can handle non-stationarity through differencing.

SARIMA (Seasonal ARIMA):

-
    - An extension of ARIMA that includes seasonal components.

2. Regression Models:

Linear Regression:

Assumes a linear relationship between predictors and the target variable.

- Can incorporate multiple external features.

Ridge and Lasso Regression:

-
    - Variants of linear regression that include regularization to prevent overfitting, especially when there are many features.

3. Tree-Based Models:

- **Decision Trees:**
  - Can capture non-linear relationships and interactions between variables.
  - Provides a clear visualization of decision rules.
- **Random Forest:**
  - An ensemble of decision trees which reduces overfitting and typically improves predictive performance.

4. Neural Networks:

- **Feed-forward Neural Networks:**
  - Can capture complex non-linear relationships.
  - Requires careful tuning and potentially large datasets.
- **Long Short-Term Memory (LSTM) Networks:**
  - A type of Recurrent Neural Network (RNN) suitable for time series data.
  - Can remember patterns over long sequences and is well-suited for multivariate time series data.

5. Hybrid Models:

- **ARIMA + Neural Network or ARIMA + XGBoost:**
  - Combine the strengths of classical time series models with the power of machine learning models.

5) .Model Training:

Model training for electricity price prediction involves taking the preprocessed data and the chosen model (or models) to learn the underlying patterns within the data. Here's a structured approach to model training for this problem:

1. Splitting the Data:

- **Time Series Split:** Because this is a time series problem, you can't randomly split data. Use an initial period for training and a later period for validation/testing.
-

2. Model Initialization:

- Choose hyperparameters for your model. For instance, if you're using ARIMA, you'll need to choose values for �p, �d, and �q.

3. Model Training:

- Feed the training data into the model, allowing it to learn the relationships within the data.
-

4. Model Validation:

- Use your trained model to make predictions on the validation set.
-
- Calculate error metrics to assess how well the model is performing. Common metrics for regression problems like electricity price prediction include:
-
    - Mean Absolute Error (MAE)
    - Root Mean Square Error (RMSE)
    - Mean Absolute Percentage Error (MAPE)

5.Final Model Training:

- Once you're satisfied with the model's performance on the validation set, train the model on the entire dataset (both training and validation) to make it ready for deployment and future predictions.

Evaluation:

Evaluation is a critical phase in the model development process, especially for a problem like electricity price prediction. Proper evaluation not only measures the performance of your model but also gives insights for potential improvements. Here's how to approach evaluation for this problem statement:

1. Choose Evaluation Metrics:

For regression problems like electricity price prediction, the following metrics are commonly used:

- **Mean Absolute Error (MAE):** Represents the average of the absolute differences between the predicted and actual values. It gives a direct idea of how much, on average, the predictions are off by.
-
-

- **Root Mean Square Error (RMSE):** Similar to MAE but penalizes large errors more heavily.
-
-
- **Mean Absolute Percentage Error (MAPE):** Represents the error as a percentage, which can be useful for understanding relative errors.
-
-
- **R-squared:** A statistical measure representing the proportion of the variance in the dependent variable that is predictable from the independent variables.
-

2. Use a Validation Set:

- Use a separate set of data (that the model hasn't seen) to evaluate performance. Given the time series nature of electricity prices, it's essential to use a chronological split.

3. Time Series Cross-Validation:

-

Due to the temporal structure of the data, traditional cross-validation techniques might not be suitable.

-

Conclusion:

In conclusion, while challenges abound, the benefits of accurate electricity price prediction are manifold. Harnessing the power of modern data science and machine learning techniques, combined with domain expertise and continuous iteration, will pave the way for robust and efficient solutions that cater to the evolving needs of the energy sector.