## **Electricity Prices Prediction**

**PROBLEM DEFNITION:**

 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.

**DESIGN THINKING:**

**Data Source:**

* Leveraging a dataset encompassing historical electricity prices, complemented by key variables like date, demand, supply, weather conditions, and economic indicators, is pivotal for in-depth analysis and decision-making within the energy sector. The dataset's core is historical electricity prices, spanning an extended timeframe, enabling the identification of pricing trends and fluctuations. Timestamps provide valuable time-series insights, revealing patterns, seasonality, and correlations with other factors. Crucial factors include electricity demand, reflecting consumption variations, and supply data, which elucidates the balance between demand and generation capacity. Weather conditions, such as temperature, humidity, and solar radiation, impact energy production and consumption during extreme weather events.
* Economic indicators shed light on electricity demand from different sectors, while regulatory changes and renewable energy generation information offer insights into market dynamics. Fuel prices, market data, grid events, and environmental factors complete the dataset, allowing stakeholders to make data-driven decisions related to pricing, supply and demand forecasting, resource optimization, and policy development. Such analysis enhances energy system efficiency, sustainability, and reliability while ensuring cost-effective electricity provision to consumers.

**Data Preprocessing:**

* Data preprocessing is a fundamental step in readying a dataset for analysis and modeling, particularly in contexts like historical electricity prices and relevant factors. Key aspects of data preprocessing include handling missing values by imputation or removal, addressing outliers through truncation or transformation, and ensuring feature scaling to put numerical attributes on a consistent scale. Conversion of categorical features into numerical representations is crucial, with methods such as one-hot encoding, label encoding, or target encoding being applied as needed. Data transformations like logarithmic scaling may be necessary for achieving a more suitable data distribution.
* Feature selection is employed to retain relevant attributes while removing unnecessary ones, reducing dimensionality. Data splitting is vital for training and evaluating models. In the case of time-series data like historical electricity prices, additional considerations include resampling, lag creation, and rolling window statistics to capture temporal patterns accurately. Comprehensive documentation of preprocessing steps ensures transparency and reproducibility in analysis, enhancing the dataset's quality for effective modeling and decision-making.

**Feature Engineering:**

* Feature engineering, a crucial aspect of machine learning, involves crafting or adapting features to bolster a model's predictive capabilities. In the realm of electricity pricing and demand forecasting, specialized feature creation can greatly enhance model performance. Time-based features, such as day of the week, month, or holidays, offer insights into temporal patterns and seasonal fluctuations. Incorporating lagged variables, like historical electricity prices or demand levels, enables the model to capture past trends. Moving averages can smooth data noise and reveal underlying patterns, while the integration of historical weather data helps account for climate-related influences on electricity usage.
* Economic indicators, such as GDP growth or inflation rates, can provide a broader context for demand fluctuations. Flags denoting special events allow for adjustments during significant occurrences. Interaction terms and custom aggregations unveil intricate relationships, and domain-specific features, tailored to the energy sector, add valuable insights. Effective feature engineering is a blend of data science expertise and domain knowledge, ultimately empowering models to uncover intricate patterns within electricity data, leading to more precise forecasts and informed decision-making in the energy industry.

**Evaluation:**

* Assessing the performance of a time series forecasting model is crucial to ensure its accuracy and reliability in predicting future values. Several established metrics are commonly used for this purpose, Mean Absolute Error (MAE) computes the average absolute differences between predicted and actual values, providing a straightforward measure of forecasting accuracy. Root Mean Squared Error (RMSE) is similar but penalizes larger errors more heavily. Lower MAE and RMSE values indicate superior model performance. Mean Absolute Percentage Error (MAPE) expresses forecast error as a percentage of actual values, offering a relative accuracy measure. Symmetric Mean Absolute Percentage Error (sMAPE) addresses MAPE's limitations.
* Other evaluation tools include assessing forecast bias, autocorrelation of forecast errors, and the Ljung-Box test to detect remaining patterns or serial correlation. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) aid in model selection and comparison. Visual inspection through predicted vs. actual plots complements quantitative metrics. Employing a combination of metrics aligns with specific forecasting objectives and business goals, ensuring a comprehensive evaluation of model performance.