Electricity Prices Prediction

Problem Statement:

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

Design thinking process:

- Utilize a dataset containing historical electricity prices and relevant factors like date, demand, supply, weather conditions, and economic indicators.
- Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.
- Create additional features that could enhance the predictive power of the model, such as time-based features and lagged variables.
- Choose suitable time series forecasting algorithms (e.g., ARIMA, LSTM) for predicting future electricity prices.
- Train the selected model using the preprocessed data.
- Evaluate the model's performance using appropriate time series forecasting metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

Data preprocessing steps:

- Gathering historical data on electricity prices. This data can be obtained from various sources, such as utility companies, energy market exchanges, or public datasets.
- Cleaning the data to remove any inconsistencies or errors. This may involve handling missing values, dealing with outliers, and correcting data formatting issues.
- Electricity price data is often collected at irregular intervals. Resample the data to a consistent time interval, such as hourly or daily, to facilitate analysis and modeling.
- Creating relevant features that can help improve the predictive power of model.
- Weather conditions like temperature, humidity, and wind speed can influence electricity demand and supply.
- Factors like GDP growth, inflation rates, and industrial production can impact electricity prices.
- Information on electricity generation, consumption, and grid congestion can be valuable.

Feature Extraction Techniques:

- Decomposing the time series data into its trend, seasonality, and residual components using techniques like Seasonal Decomposition of Time Series (STL) or moving averages.
- Using feature selection techniques to identify the most important features that contribute to price prediction. Techniques like recursive feature elimination or feature importance from tree-based models can help with this.
- Evaluating the performance of your model on the test dataset using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), or Mean Absolute Percentage Error (MAPE).

Machine learning algorithm used:

Random Forest Regressor:

- Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. Each tree is built independently, and the final prediction is an average or weighted combination of the individual tree predictions.
- Random Forest is a non-parametric algorithm, meaning it doesn't make strong assumptions about the underlying data distribution. This makes it versatile and suitable for various types of data, including electricity price time series data.
- Random Forest can provide information about feature importance, which helps you identify which features have the most significant impact on electricity price prediction. This can be valuable for feature selection and understanding the driving factors behind price fluctuations.
- Electricity price data often exhibits complex, non-linear patterns. Random Forest can capture these non-linear relationships between features and the target variable, making it suitable for modeling such data.
- Random Forest is less prone to overfitting compared to individual decision trees, thanks to techniques like bagging (Bootstrap Aggregating) and feature randomization. These techniques enhance model generalization.
- Random Forest is robust to outliers in the data because it averages predictions from multiple trees, reducing the impact of extreme values.

Model Training:

- Choosing an appropriate model for electricity price prediction. Common choices include regression models (e.g., Linear Regression), time series models (e.g., ARIMA), machine learning models (e.g., Random Forest, Gradient Boosting), and deep learning models (e.g., Recurrent Neural Networks or Long Short-Term Memory networks, LSTM). In this project we have chosen **RandomForestRegressor.**
- Training the selected model on the training dataset. The model learns to make predictions by finding patterns and relationships in the historical data.

- Assessing the model's performance using appropriate regression metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R2).

Innovative approach:

- Decomposing time series data into its trend, seasonality, and residual components using advanced techniques like Seasonal Decomposition of Time Series (STL) or Singular Spectrum Analysis (SSA). This helps in capturing and modeling the underlying patterns and fluctuations in electricity prices.
- Combining traditional time series models with machine learning models like Random Forest or Gradient Boosting for better accuracy.