Synopsis: Load Forecasting and Missing Value Imputation using Random Forest Regression

This project explores the use of a Random Forest Regression model to predict electricity load and fill in missing data points within historical load records. Accurate load forecasting is critical for grid operators to ensure reliable power supply, optimize resource allocation, and minimize operational costs. Furthermore, complete and accurate historical load data is essential for various analyses, including demand trend identification, capacity planning, and anomaly detection.

Approach:

The project involved several key steps:

1. **Data Preprocessing:**
    - Acquired historical load data, including time-stamped load values, weather information (temperature, humidity, wind speed), and other relevant factors.
    - Cleaned the data by handling missing values (initially using interpolation) and addressing outliers.
    - Engineered relevant features such as time-based features and weather-related features.
    - Split the dataset into training, testing sets.
2. **Model Training:**
    - Trained a Random Forest Regression model on the prepared training data. Random Forest was chosen for its ability to handle non-linear relationships, capture feature interactions, and provide robust predictions.
3. **Model Evaluation:**
    - Evaluated the model's performance on the validation set using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).
4. **Load Forecasting:**
    - Utilized the trained model to predict future load values for a specific time horizon.
5. **Missing Value Imputation:**
    - Employed the model to fill in missing values in the historical load data by predicting the load value based on the available features for each missing data point.
    -

Insights:

- **Feature Engineering:** The careful selection and engineering of features significantly influenced model performance. Time-based features (hour, day of week) proved crucial in capturing daily and weekly load patterns. Weather-related features, particularly temperature, demonstrated a strong correlation with electricity demand.
- **Model Performance:** Random Forest Regression exhibited good performance in load forecasting, effectively capturing the non-linear relationships between features and load. However, the specific performance varied depending on data quality, load pattern complexity, and the accuracy of feature engineering.
- **Model Interpretability:** While Random Forest models can be complex, techniques like feature importance analysis provided insights into which features had the greatest influence on load predictions, aiding in model understanding and identifying areas for potential improvement.

Conclusions:

This project successfully demonstrated the application of a Random Forest Regression model for both load forecasting and missing value imputation in electricity load data. The results emphasized the importance of data preprocessing, feature engineering, and appropriate model selection in achieving accurate and reliable predictions.

Future Directions:

- **Hyperparameter Tuning:** Optimizing model hyperparameters (e.g., number of trees, tree depth) to enhance performance.
- **Incorporating External Factors:** Including additional relevant factors such as economic indicators, special events, and social media trends.
- **Exploring Alternative Models:** Investigating other machine learning models, such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), or more advanced time series models like Long Short-Term Memory (LSTM) networks.
- **Ensemble Methods:** Combining predictions from multiple models (e.g., bagging, boosting) to improve overall accuracy and robustness.