**Documentation: Smart Grid Optimization**

**Problem Statement:**

Aims to develop innovative machine learning solutions to optimize energy generation, distribution, and consumption within a smart grid environment. Participants are tasked with predicting total power generation, distributing generated power among consumption nodes, assessing grid stability, and providing valuable insights into grid performance trends.

**Approach:**

1. Data Preparation:

- Merge 20 files containing data on air temperature, pressure, wind speed, and power generated.

- Combine independent variables for the first three months of 2024.

- Prepare a final dataset with timestamps at hourly granularity.

2. Total Power Generation Prediction:

- Random forest classifier used to classify the data after removing the outliers and handling the missing values

- Data is train\_test\_split to generate the confusion matrix

- Power Generated (MW): power\_gen\_1, power\_gen\_2, power\_gen\_3 ,the total power

generated is distributed into each node as 20%, 45% & 35% which is calculated by generating the the total power consumption and dividing into 3 node upon each percentage given and displayed in the unit consumption column

* Classification matrix and confusion matrix is developed upon the validation of the model and precision,recall ,f1 score and support is calculated.
* The percentage stability and unstability of each node is calculated nxt

3. Power Distribution to Consumption Nodes:

- Allocate generated power to three consumption nodes based on specified percentages.

- Node 1: 20% of total power

- Node 2: 45% of total power

- Node 3: 35% of total power

Outliers removed by z score 406 outliers were found in wind prediction

Xgb regression

Validator score : 99.99

4. Grid Stability Assessment:

- Use the existing and newly prepared dataset to determine grid stability.

- Classify stability as 'stable' or 'unstable' based on predefined criteria.

- Employ classification models like Logistic Regression, Decision Trees, or Support Vector Machines.

5. Insights and Reporting:

- Analyze the output to determine stability trends over the three-month period.

- Identify peak hours of instability and patterns in unstable conditions.

- Provide insights to inform grid operators about potential vulnerabilities and proactive measures.

**Implementation Ideas:**

1. Data Preparation:

- Load and concatenate data from 20 files into a single dataframe.

- Merge independent variables for Jan, Feb, and March 2024 with the final dataset.

2. Total Power Generation Prediction:

- Split data into training and validation sets.

- Train various regression models on historical data.

- Select the best-performing model based on validation metrics such as RMSE or MAE.

- Predict total power generation for the first three months of 2024.

3. Power Distribution:

- Calculate power distribution percentages for each node.

- Allocate generated power accordingly to nodes.

4. Grid Stability Assessment:

- Split data into features and target variables.

- Train classification models on historical data to predict grid stability.

- Evaluate models using metrics like accuracy, precision, and recall.

- Classify grid stability for the three-month period.

5. Reporting:

- Generate visualizations to present insights on stability trends and peak hours of instability.

- Prepare a comprehensive report summarizing model performance, stability analysis, and actionable insights.

**Best Models Selection:**

Total Power Generation Prediction:

- : Offers high predictive accuracy and robustness against overfitting, suitable for complex relationships in the data.

Grid Stability Assessment:

- Random Forest Classifier: Effective in handling high-dimensional data and capturing non-linear relationships, providing reliable stability predictions.