FROYO TECHNOLOGIES
DATA AI FOR BUSINESS

# Level-2-Data Analytics and Predictive Modeling Project Documentation

**Project Title:**

**Energy Consumption Analysis**

**Team Members:**

**Soumajit Mandal**

**Project Overview:**

The goal of this project is to analyze energy consumption data to identify patterns, optimize usage, reduce costs, and enhance sustainability. By leveraging data analytics, the project aims to provide actionable insights for energy management and conservation strategies. The project addresses the problem of inefficient energy usage and aims to offer solutions for better energy management.

## 1. Introduction

- **Purpose:** The purpose of this project is to analyze household energy consumption patterns to optimize energy usage, reduce costs, and promote sustainability.
- **Scope:** The scope includes data analysis of household energy consumption patterns, anomaly detection, predictive modeling for forecasting future consumption, and recommendations for energy efficiency strategies.

## 2. Data Collection

- **Data Sources:** The dataset is sourced from the UCI Machine Learning Repository, which includes household power consumption data from various electrical appliances.
- **Data Acquisition:** The data was retrieved from the UCI repository using automated download methods, ensuring data integrity and completeness.

## 3. Data Preprocessing

- **Data Cleaning:** Missing values were handled by filling them with the mean of each column. Duplicate entries were checked and corrected if necessary.
- **Data Transformation:** Numerical data (such as Global_active_power, Global_reactive_power, Voltage, Global_intensity) were normalized using MinMaxScaler to standardize the scale across different features.
- **Data Integration:** The dataset was integrated into a unified format suitable for analysis, ensuring consistency across different data sources.

## 4. Exploratory Data Analysis (EDA)

- **Descriptive Statistics:** Summary statistics provided insights into the central tendencies and variability of energy consumption across different time periods.
- **Visualization:** Visualizations included histograms to depict the distribution of Global_active_power, time series plots to visualize trends in energy consumption over time, and correlation matrices to analyze relationships between variables.
- **Correlations:** The correlation analysis highlighted significant relationships between energy consumption and factors such as voltage, intensity, and reactive power.

## 5. Feature Engineering

- **Feature Selection:** Relevant features (Global_reactive_power, Voltage, Global_intensity) were selected based on their impact on Global_active_power.
- **Feature Creation:** No new features were created in this phase, as the focus was on utilizing existing features to predict energy consumption patterns effectively.

## 6. Model Development

- **Model Selection:** A Linear Regression model was chosen for predicting Global_active_power, and an Isolation Forest model was used for anomaly detection.
- **Model Training:** Models were trained on a subset of the data, with careful consideration of hyperparameters to ensure optimal performance.
- **Model Evaluation:** Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared were used to assess the performance of both models.

## 7. Model Interpretation

- **Feature Importance:** In the Linear Regression model, Global_intensity and Voltage were found to be the most significant features in predicting Global_active_power.
- **Model Insights:** The models provided insights into peak energy consumption periods, aiding in the identification of strategies for load shifting and energy efficiency improvements.

## 8. Model Deployment

- **Deployment Plan:** The models were deployed using local environments for initial testing, with plans to integrate them into a cloud-based platform for scalability.
- **Monitoring and Maintenance:** Post-deployment, the models will be monitored regularly to ensure their accuracy and effectiveness in predicting energy consumption patterns.

## 9. Conclusion

- **Summary:** The project successfully analyzed household energy consumption data, identified patterns, and provided recommendations for optimizing energy usage and promoting sustainability.
- **Challenges:** Challenges included handling missing data and optimizing model performance, which were addressed through rigorous data preprocessing and model tuning.
- **Future Work:** Future research could explore incorporating weather data and smart meter analytics to enhance predictive accuracy and further reduce energy consumption.

## 10. Recommendations:

### 1. Load Shifting Strategies

- **Peak Hours Identification:** The analysis identified peak consumption hours during specific times of the day. To optimize energy usage, it is recommended to shift high-energy-consuming activities to off-peak hours when the demand is lower.

### 2. Energy Efficiency Improvements

- **Upgrade to Energy-Efficient Appliances:** Encourage the use of energy-efficient appliances, such as LED lighting, ENERGY STAR-rated devices, and smart thermostats, which consume less power while maintaining performance.

### 3. Renewable Energy Integration

- **Solar Panel Installation:** Promote the installation of solar panels to harness renewable energy, which can significantly reduce reliance on non-renewable energy sources and lower overall energy costs.

### 4. Smart Home Technology

- **Smart Meters:** Deploy smart meters that provide real-time data on energy usage, helping consumers understand their consumption patterns and identify opportunities for savings.