# Predictive Trend Analysis of Household Energy Consumption

## 1. Problem Statement

Households and utilities often lack predictive foresight, making demand spikes, billing shocks, and grid strain harder to manage. This project tests whether sequence model approach using LSTM or GRU can reliably predict short-term consumption trends using past loads and exogenous variables (e.g., temperature), and whether such forecasts can inform actionable insights.

## 2. Activity Overview

The project aims to develop a predictive model for household electricity consumption using historical data and advanced machine learning techniques.

The primary activities include:

- Data Acquisition & Cleaning: Utilize the household power consumption dataset (2006–2010) containing variables such as Global Active Power, Voltage, and Sub-metering values.
- Exploratory Data Analysis (EDA): Identify consumption patterns, seasonal trends, and anomalies.
- Feature Engineering: Create time-based features (year, month, day, weekday), lag features, and rolling statistics.
- Model Development: Implement an LSTM/GRU time series forecasting model.
- Evaluation & Visualization: Assess model performance using MAE and RMSE; visualize predictions and trends.
- Deployment: Visualizations for real-time monitoring and forecasting.

## 3. Scope

### In-Scope

- Data preparation and ingestion.
- LSTM/GRU model development and evaluation.
- Baseline comparator model.
- Data Insights.

## 4. Methodology

The proposed approach will utilize Long Short-Term Memory (LSTM) and Gated Recurring Unit (GRU) approach due to its strength in modeling temporal dependencies and handling sequential data, which is essential for forecasting household power consumption. LSTM networks are particularly effective in capturing long-term patterns and trends in time series data, making them suitable for this application.

The input to the LSTM/GRU model will consist of time-stamped power consumption readings, along with engineered features such as day of the week, hour of the day, and lagged consumption values. These features help the model learn seasonal and daily consumption patterns. The output will be a forecast of future power consumption values over a specified horizon.
The model architecture will include multiple LSTM/GRU layers followed by dense layers to refine the output. Dropout layers will be incorporated to prevent overfitting. The final layer will produce a single or multi-step forecast depending on the configuration.

### Data Preprocessing

The preprocessing pipeline will include handling missing values using interpolation or forward-fill techniques, normalization of numerical features to ensure consistent scaling, and feature extraction such as time-based features and rolling averages. Seasonal decomposition may also be applied to isolate trend and seasonal components.

### Evaluation Strategy

Model performance will be evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE). A validation strategy involving time-based train-test splits will be employed to ensure robust evaluation.

### LSTM/GRU Modeling Details

- Input window: 7–14 days of historical data.
- Output window: 24 hours (primary horizon).
- Features: lagged consumption, weather variables, calendar flags.
- Model: LSTM/GRU layers with dropout and dense output head.
- Evaluation: Rolling-origin back testing with MAE, RMSE.

## 5. Architecture Diagram

Diagram: Predictive System Architecture


## 6. Deliverables

- Google Collab notebook for data processing, EDA and time series prediction models.
- Visualizations illustrating consumption patterns, model evaluation metrics, and predicted trends.

## 7. Success Criteria

- Consistent performance across rolling back tests.
- Clear visualization of forecasts and drivers.