Household Energy Consumption Prediction

1. Problem Understanding

Overview

The aim of this project is to **predict household electricity consumption** using historical smart meter data combined with weather information. Predicting energy usage accurately can help consumers and utility providers optimize power usage, reduce costs, and manage energy distribution more efficiently.

Problem Motivation

Energy consumption patterns are influenced not just by past usage but also influenced by external conditions such as temperature, humidity, and seasonal variations. Simple threshold-based methods often fail to capture these complex patterns, so machine learning becomes essential for **accurate forecasting**.

Real-World Implications

Accurate energy forecasting enables:

- Smart grid load balancing
- Reduced electricity costs for consumers
- Predictive energy planning for utility companies
- Integration with IoT systems in smart homes

Formal Problem

Given time-indexed energy meter readings and weather data, the model should:

- **Input:** Historical energy and weather features
- **Output:** Predicted future energy consumption
- **Goal:** Minimize prediction error while capturing patterns arising from time and weather influence

2. Model Pipeline Description

Data Sources

Your solution integrates two data sources:

1. **Household Power Consumption Dataset** — smart meter readings with measurements like Global Active Power, Voltage, Sub-metering (Energy usage) over time.
2. **Weather Data** — external conditions such as temperature, precipitation, wind speed, and pressure that influence energy use.

Data Preprocessing

Key steps performed in the notebook:

- **Cleaning:** Handling missing values, ensuring datetime formatting
- **Datetime Feature Engineering:** Extracting units like day, hour, month to capture periodic trends
- **Merging:** Align energy and weather datasets into a combined time series
- Feature scaling as needed for machine learning inputs.

Feature Engineering

You generated:

- Lag features (previous consumption)
- Seasonal indicators (month, day, hour)
- Combined weather + power usage features
  These help the model understand patterns across time and environment.

Model Used

A **Linear Regression model** is trained on the engineered dataset.
Although simple, it performs well due to strong linear trends in the data.

Training Procedure

- Data split into training and test sets
- Features and labels separated
- Model trained to minimize error between predictions and actual consumption

3. Results & Metrics

Evaluation Metrics

The notebook reports these performance metrics after training:

Metric	Description
RMSE	Root Mean Square Error — measures average prediction error
MAE	Mean Absolute Error — average absolute difference
R-Squared	Proportion of variance explained by model

Your model achieves a **very high R-Squared value (~99.86%)** — indicating strong alignment with actual energy usage trends.

Interpretation

- **High R-Squared** suggests the model fits the training data extremely well.
- Lower RMSE and MAE indicate tight prediction boundaries and minimal average errors.
- The combination of temporal and weather features provides additional predictive power.

4. Code (GitHub / Colab)

🔗 https://github.com/Hymavathi07/Cerevyn_Solutions

Notebook

Main work is in the Jupyter Notebook:

📄 Household_energy_consumption_prediction.ipynb — contains all experiments from loading data to model evaluation.

Key Files

Filename	Purpose
Household_energy_consumption_prediction.ipynb	End-to-end prediction pipeline
household_power_consumption.csv	Energy dataset
sceaux_weather_data.csv	Weather dataset

Filename	Purpose
requirements.txt	Environment setup
README.md	Project overview

Installation

To run locally or on Colab:

git clone https://github.com/Hymavathi07/Cerevyn_Solutions.git

pip install -r requirements.txt

🧠 **Notebook Structure**

1. Import libraries
2. Load data
3. Preprocessing & feature engineering
4. Model training
5. Evaluation & visualization
6. Interpretation of results

Conclusion

- You have successfully built a **predictive model** that uses both historical and external weather data.
- The model demonstrates **excellent performance**, especially in variance explanation (R-Squared ~99.86%).
- This project helps understand energy usage patterns and can be extended to real-time systems.

Future Scope

- Integration with **IoT & Smart Devices** for real-time predictive input.
- Use of **advanced models** (like Random Forests, LSTM) for improved pattern extraction.
- Support for multiple households or region-wide forecasting.