Final Project Progress Report

Power Load Forecasting

Xinyuan Huang

November 12, 2025

Project scope update

This project investigates short-term household power load forecasting using time-series and machine learning models. The work initially considered ARIMA as a baseline, but the seasonal structure and periodic patterns in the data motivated a switch to **SARIMA**.

The final comparison focused on four models: **SARIMA**, **ETS** (Exponential Smoothing), **XGBoost**, and **LSTM**. Across consistent train/test splits and common features, the results showed that **LSTM achieved the best accuracy**, **XGBoost ranked second**, while **ETS** was stable and competitive on seasonal signals but less flexible, and **SARIMA** captured seasonality explicitly yet underfit nonlinear effects.

Data sources

Data were obtained from the following sources:

- **UCI Household Power Consumption (2006–2010)** — provides minute-level household power readings, aggregated to hourly intervals for modeling consistency. Download link: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption
- **Meteostat API** — provides hourly **temperature**, **humidity**, and **wind speed** data for **Paris**.

The datasets were merged using timestamps and resampled to hourly intervals. Missing data were handled through linear interpolation, and normalization was applied to ensure feature comparability.

Issues / difficulties

- **Model adaptability:** Classical seasonal models (SARIMA/ETS) struggled with complex nonlinear patterns; LSTM/XGBoost handled them better.
- **Data alignment:** Power and weather data used different time resolutions; careful resampling/synchronization was required.
- **Computation cost:** LSTM needed longer training time and more compute; hyperparameter tuning was nontrivial.
- **Data quality:** Occasional gaps in weather data slightly reduced usable samples and affected evaluation fairness.

Repository: https://github.com/xhuang21/power-load-forecasting