Power Load Forecasting:
A Comparative Study of SARIMA, ETS, XGBoost, and LSTM

Huang Xinyuan
DSCI 510 – Final Project

01

Project Introduction

◆ Background & Problem

Electricity demand varies significantly by hour, day, and weather conditions. These fluctuations make short-term load forecasting essential for maintaining grid stability and ensuring efficient and reliable power distribution.

◆ Why It Matters

Accurate forecasting helps prevent over- and under-supply, reduces operational costs, and supports optimized generation planning. Traditional statistical models often struggle to capture nonlinear consumption patterns found in real-world energy data.

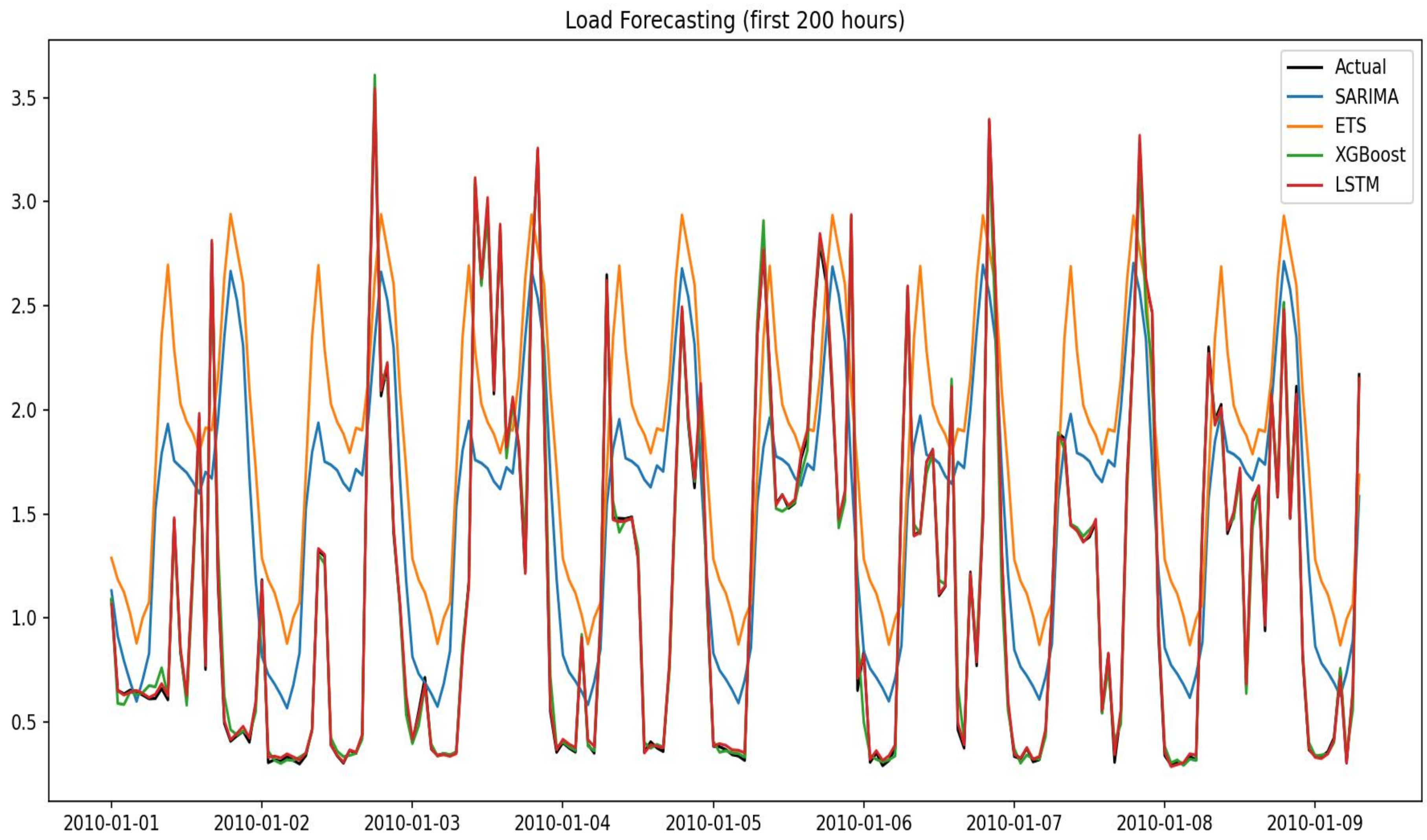
◆ Objective of This Project

This project compares four forecasting approaches—SARIMA, ETS, XGBoost, and LSTM—using historical power consumption and weather data. The goal is to identify the most accurate and robust model for short-term hourly load prediction.

Data Sources

Data Source	Description	Processing Approach	Purpose
Raw Household Power Consumption (UCI)	1-min power measurements (2006–2010): active/reactive power, voltage, current.	Parsed datetime, removed missing/invalid entries, standardized units.	Base raw signal for hourly load construction.
Meteostat Hourly Weather (Paris)	Hourly temperature, humidity, wind speed from 2006–2010.	API fetch → clean → standardize → merge by datetime.	Exogenous weather features.
Constructed Hourly Load	Aggregated hourly consumption from UCI raw data.	Resampled to hourly, filled gaps, aligned timestamps.	Main target variable (y) for forecasting.
Derived Load & Weather Features	Lag features (1h, 24h), rolling averages.	Applied shifting, rolling windows, normalization.	Improve XGBoost/LSTM performance.
Modeling Dataset (train/test)	Final merged dataset of load + weather + features.	Normalized and split (2006–2009 train, 2010 test).	Used to evaluate SARIMA, ETS, XGBoost, LSTM.

02 Summary of the Results



Key Observations

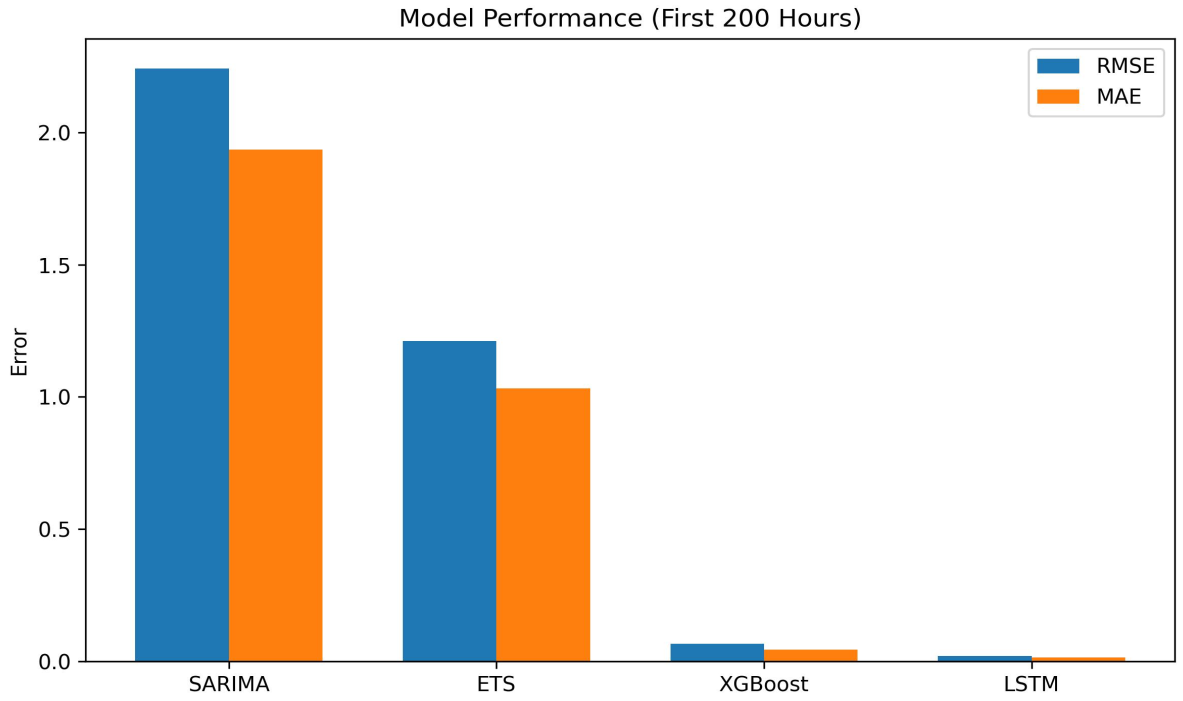
- LSTM follows the actual load curve almost perfectly, showing the best fit among all models.
- XGBoost captures the main trend well with small deviations, especially around peak hours.
- ETS tends to smooth out rapid load changes, missing sharp rises and drops.
- SARIMA struggles most significantly, with clear under- and over-estimation during highly fluctuating periods.

Interpretation

- Short-term electricity load exhibits strong nonlinear and weather-driven fluctuations.
- Only machine learning (XGBoost) and deep learning (LSTM) can effectively capture these complex variations.

Error Comparison (RMSE & MAE)

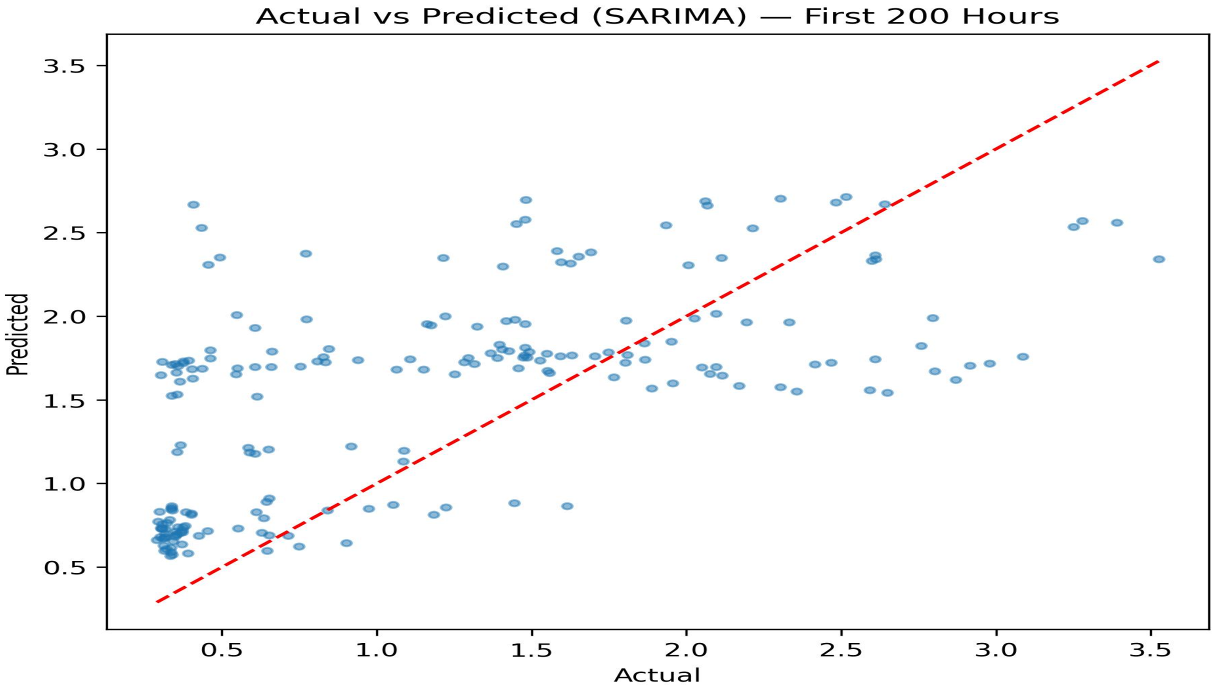
Model	MAE	RMSE
SARIMA	1.9351249833184387	2.242362329413668
ETS	1.032889010959957	1.2115953617687734
XGBoost	0.0439903588926927	0.06627018540870046
LSTM	0.015326186049787402	0.019823957453569293



Summary

- LSTM achieves the best overall performance, with the lowest MAE and RMSE among all models.
- XGBoost ranks second, showing strong and stable accuracy with small errors.
- ETS produces moderate accuracy, but its errors remain noticeably higher than those of LSTM and XGBoost.
- SARIMA performs the worst, with the largest MAE and RMSE.
- Overall, machine learning and deep learning models consistently outperform classical statistical models in short-term load forecasting.

Error Analysis — SARIMA & ETS

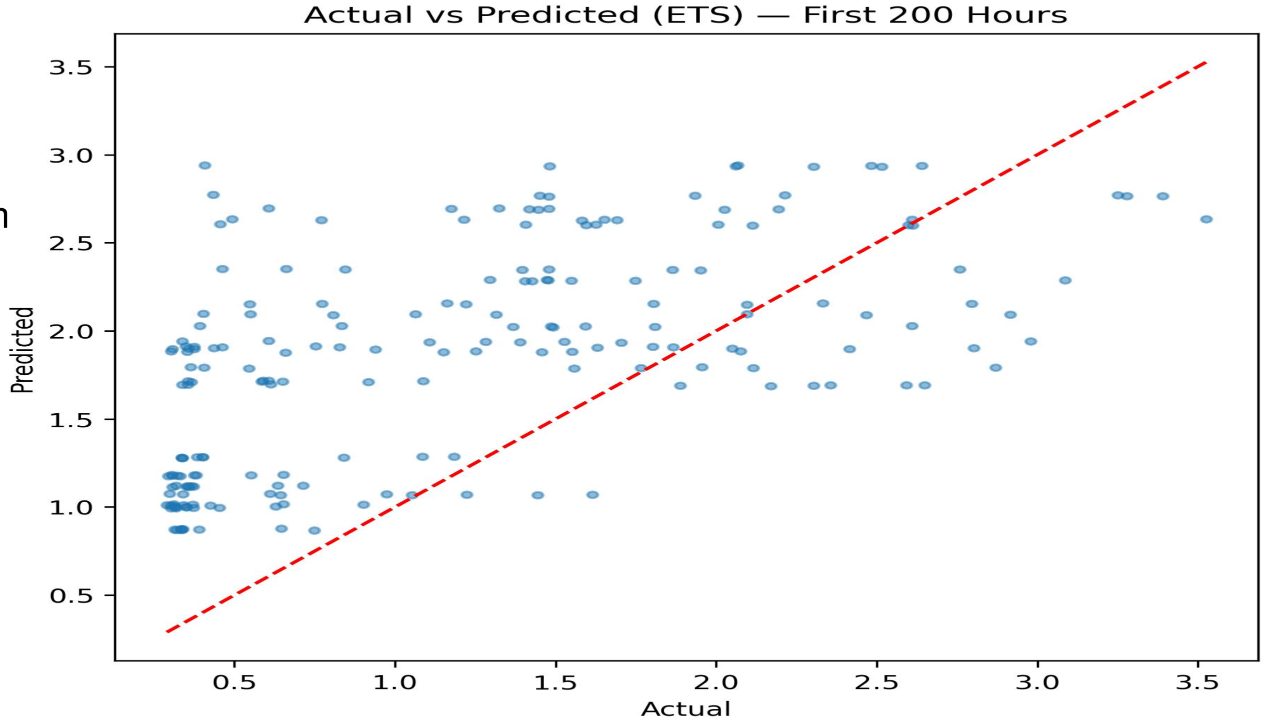


SARIMA – Key Observations

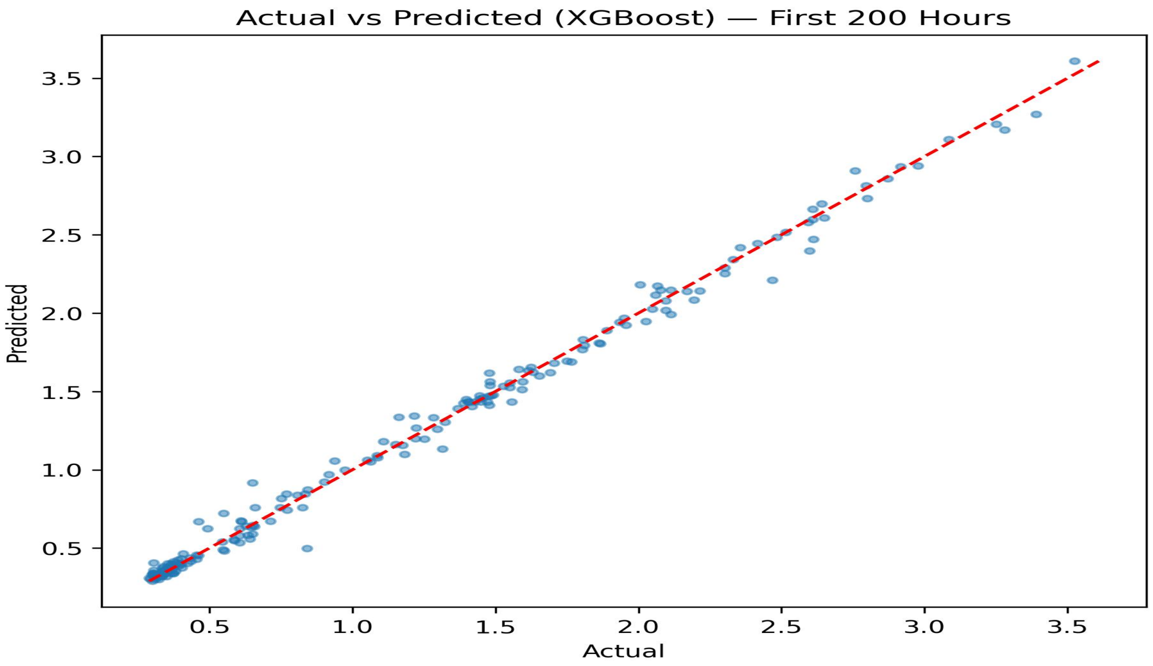
- SARIMA shows the weakest fit among all models, with points widely scattered away from the diagonal reference line.
- The model frequently exhibits large under-estimation and over-estimation, especially during rapid fluctuations in electricity load.
- SARIMA relies on assumptions of linearity and stationarity, which do not match the nonlinear, weather-driven nature of real power consumption.
- As a result, it struggles to capture sharp changes, peaks, and irregular patterns, leading to the highest MAE and RMSE in the comparison table.

ETS – Key Observations

- ETS performs better than SARIMA but still shows large deviations from the diagonal reference line.
- The model tends to smooth out rapid changes, causing missed peaks or delayed responses during fluctuating load periods.
- ETS relies on exponential smoothing, assuming the series can be represented by slowly updated trend and seasonal components.
- As a result, the model often underestimates high-demand hours and overestimates low-demand hours, revealing its limited flexibility in handling nonlinear load patterns.
- Therefore, ETS produces moderate but noticeably higher errors compared with XGBoost and LSTM.



Error Analysis —XGBoost & LSTM

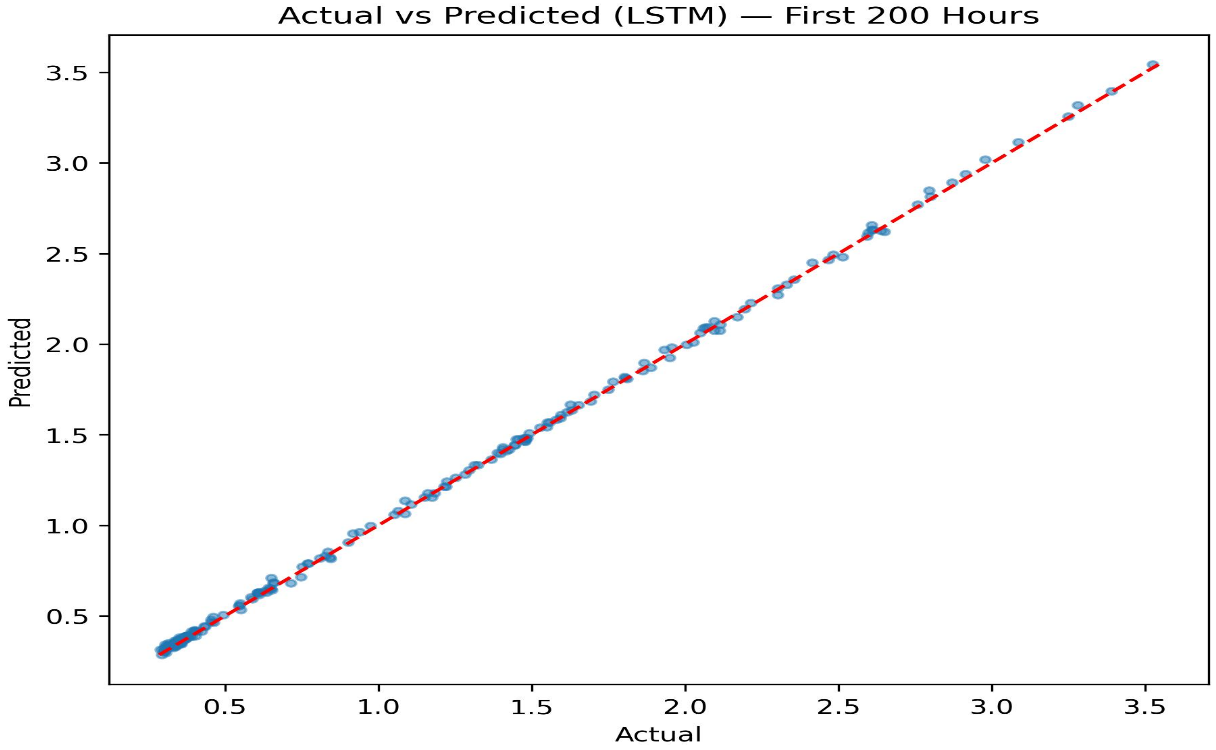


XGBoost – Key Observations

- XGBoost shows a very strong fit, with points concentrated close to the diagonal reference line.
- The model accurately captures nonlinear relationships in electricity load data due to its tree-based structure.
- It performs well during both high-demand and low-demand periods, with minimal under- or over-estimation.
- XGBoost handles interactions and nonlinear patterns effectively, allowing it to react to sudden peaks or drops that linear models cannot capture.
- As a result, XGBoost achieves significantly lower MAE and RMSE compared with SARIMA and ETS.

LSTM – Key Observations

- LSTM provides the closest alignment to the diagonal line among all models, indicating the highest prediction accuracy.
- The model captures long-term temporal dependencies through memory-cell and gating mechanisms.
- LSTM handles nonlinear, weather-driven load variations exceptionally well, outperforming both linear (SARIMA, ETS) and tree-based models (XGBoost).
- It shows minimal under-estimation or over-estimation, even during rapid fluctuations or peak hours.
- As a deep learning model, LSTM generalizes complex relationships most effectively, resulting in the lowest MAE and RMSE.



03

Challenges in This Project

•Data preprocessing: Cleaning 1-minute raw load data and converting it into reliable hourly series without losing important patterns.

•Feature alignment: Ensuring weather data and load data match perfectly in timestamp and frequency.

•Model sensitivity: Classical models require stationarity; ML/DL models require careful tuning to avoid instability.

•Capturing nonlinear behavior: Sharp peaks and weather-driven fluctuations are difficult for traditional models to learn.

Thank You!