Phase 2: Model Evaluation Report

Author: Mantas Valantinavicius

Date: Auto-generated

# 1. Project Overview

This report summarizes the results of Phase 2 of the energy forecasting dashboard project. The objective was to train and compare multiple time-series forecasting models to predict hourly energy consumption. All modeling is performed using metric system inputs and outputs: energy in kilowatt-hours (kWh), temperature in degrees Celsius (°C), and time in hourly intervals.

# 2. Methodology

The synthetic dataset was enriched with temporal, cyclical, and lag-based features. Models were trained on 80% of the dataset and evaluated using RMSE, MAE, and MAPE, all expressed in kilowatt-hours (kWh). XGBoost was configured with 100 estimators and a maximum depth of 5. Prophet included daily and weekly seasonality. Linear Regression was used as a baseline model for comparison.

# 3. Model Performance Summary

|  |  |  |  |
| --- | --- | --- | --- |
| model | RMSE | MAE | MAPE |
| Prophet | 0.1036 | 0.0814 | 17.0442 |
| XGBoost | 0.0787 | 0.0595 | 12.3052 |
| Linear | 0.1125 | 0.0866 | 17.8052 |

Note: All error metrics (RMSE, MAE, MAPE) are reported in kilowatt-hours (kWh).

# 4. Actual vs Predicted Plots

Note: All vertical axes represent energy use in kilowatt-hours (kWh).

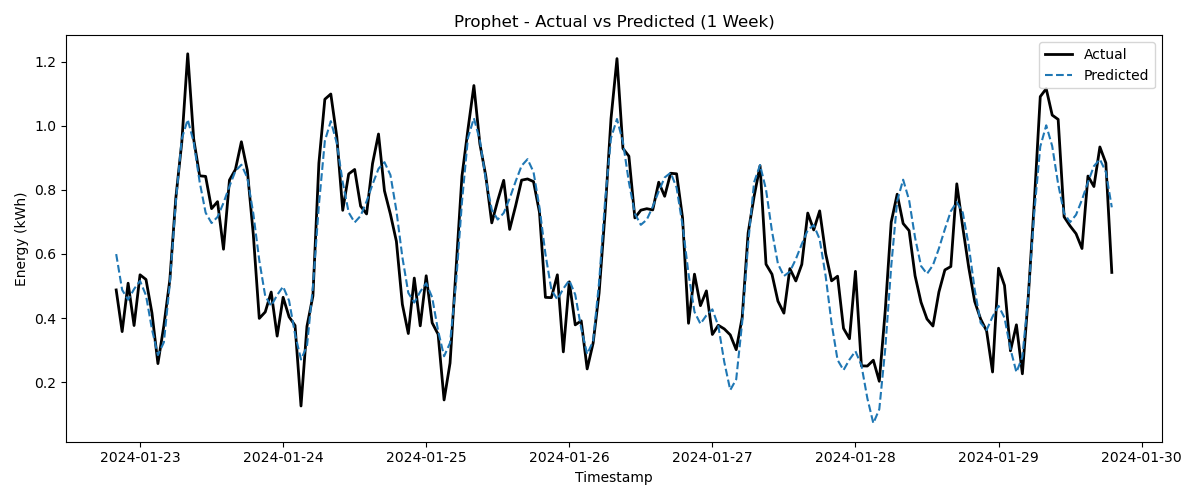


Figure: Prophet model predictions

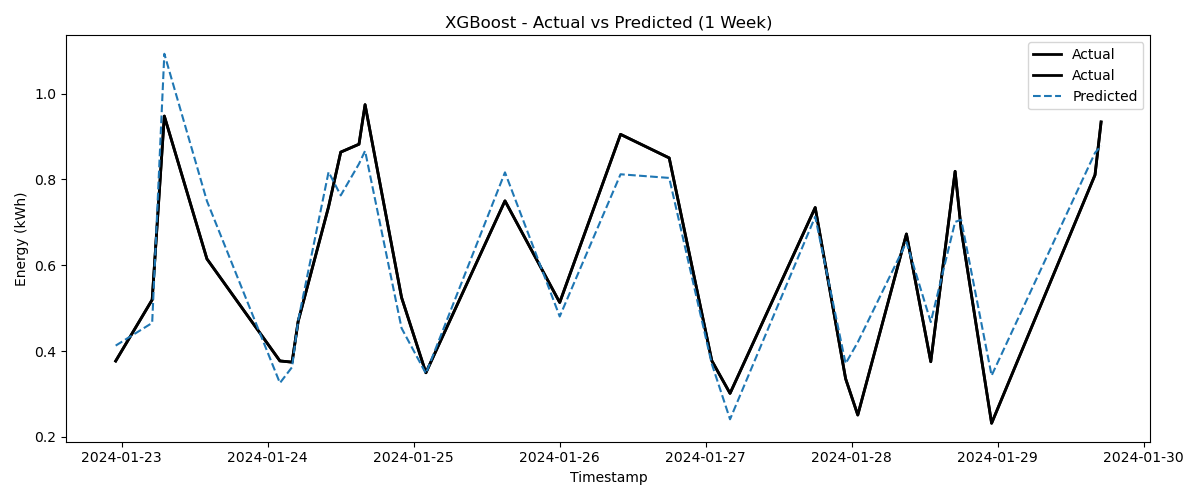


Figure: Xgboost model predictions

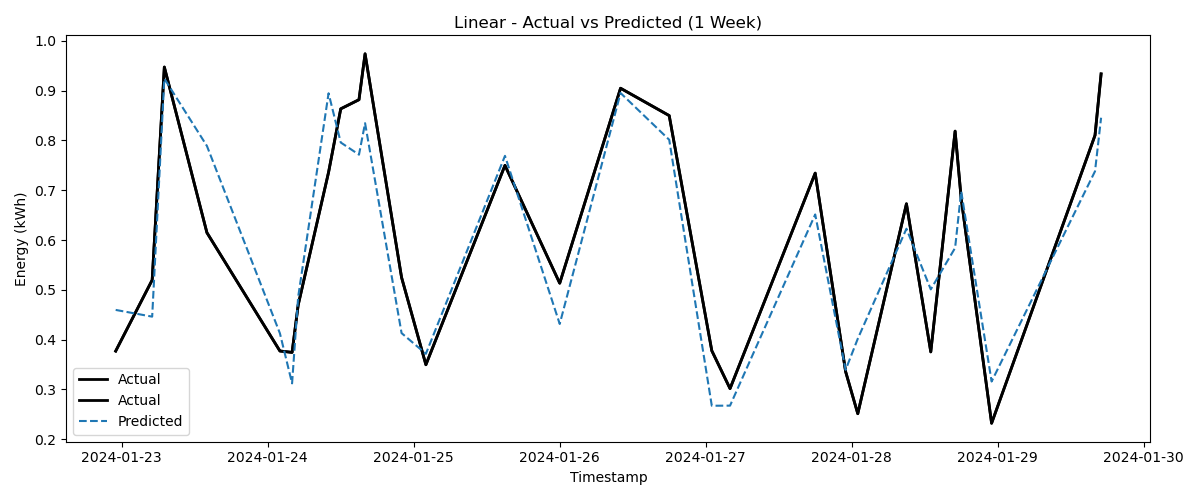


Figure: Linear model predictions

# 5. Feature Importance (XGBoost)

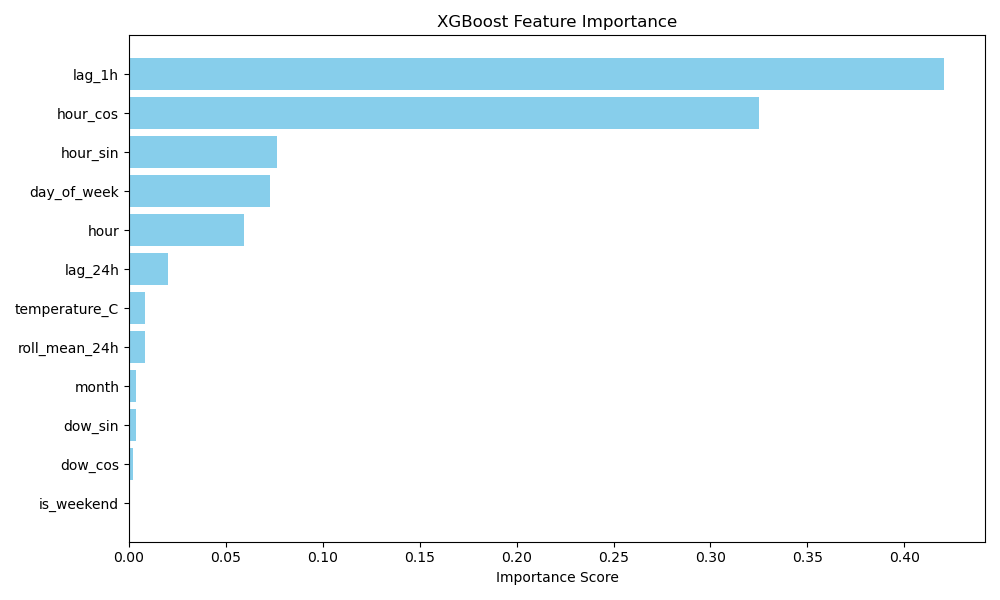


Figure: Relative importance of each feature

# 6. Feature Coefficients (Linear Regression)

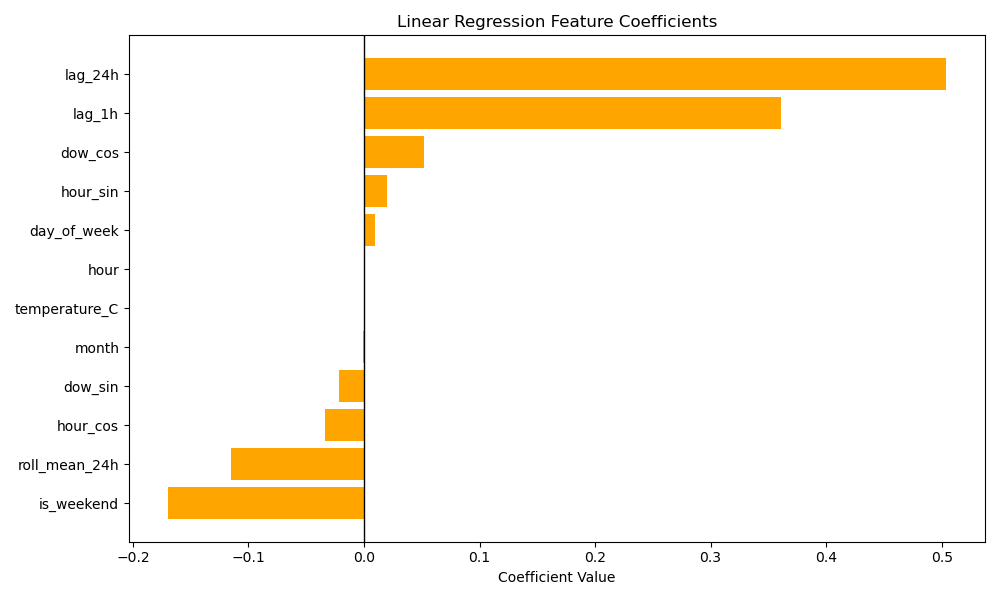


Figure: Signed influence of each input feature

# 7. Prophet Model Components

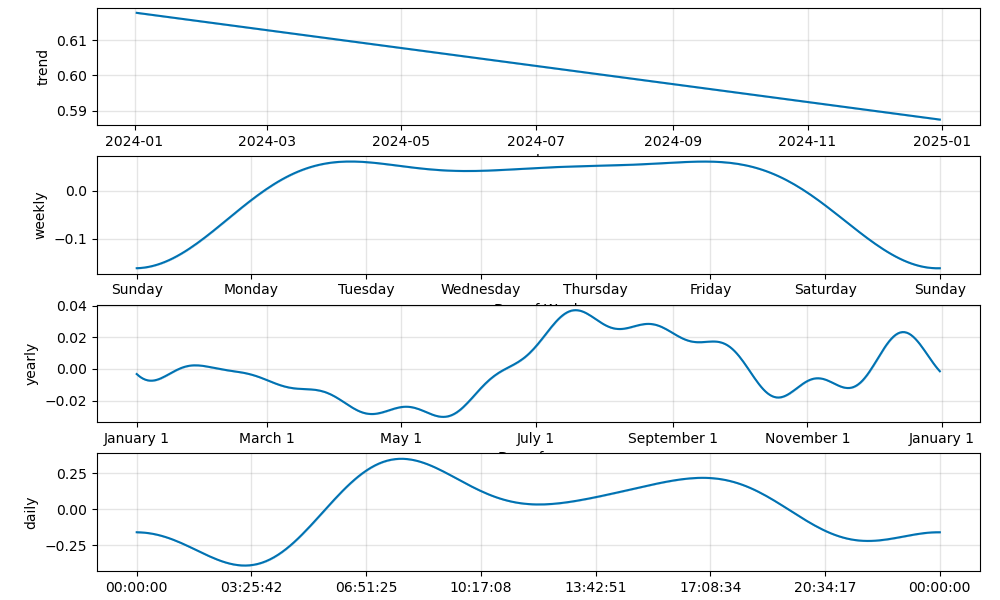


Figure: Prophet trend and daily seasonality components

The above plot illustrates the components learned by Prophet, including overall trend and repeating daily cycles. These cycles reflect typical human activity patterns, such as energy use peaks during working hours.

# 8. Conclusion

Among the evaluated models, XGBoost produced the best performance, achieving the lowest error across all metrics. Its effectiveness is attributed to strong use of time-based features and lagged values. Prophet offered strong interpretability through trend and seasonality decomposition. Linear Regression served as a useful but limited benchmark. All models were trained and assessed using metric units for real-world applicability.