



Load Forecasting for 33kV/11kV Substation

Using XGBoost & Ensemble Learning Methods

Ahmed Khalifa Alsudairy

Ahmed Mohammed Alyahya

Lina Al-badi

Aysha Alqaithi

Supervisor: Shaik Abdul Saleem

University of Technology and Applied Sciences

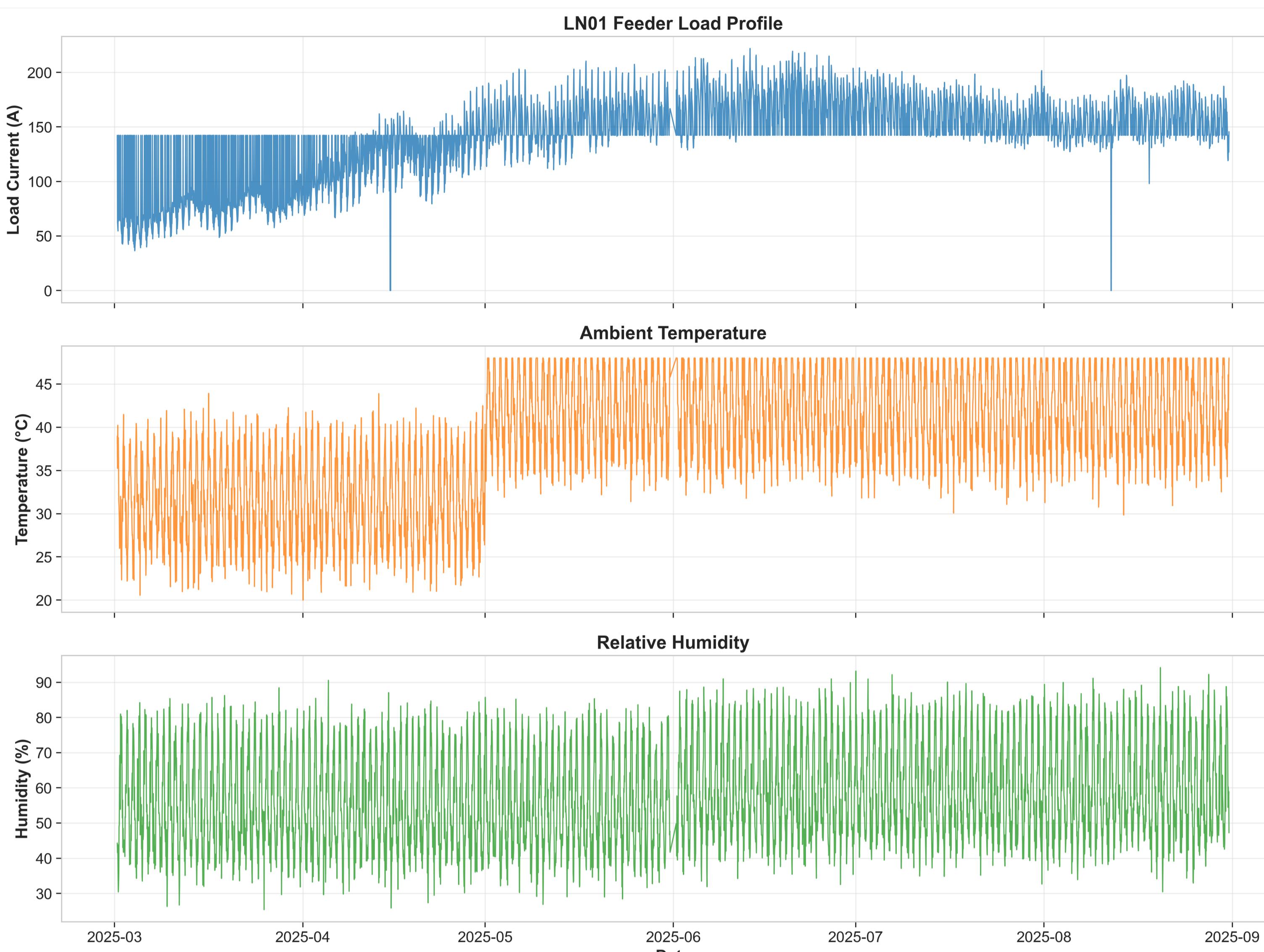
1 Introduction & Problem

The Challenge

Current manual forecasting methods fail to capture **nonlinear demand patterns**, leading to grid instability and increased operational costs.

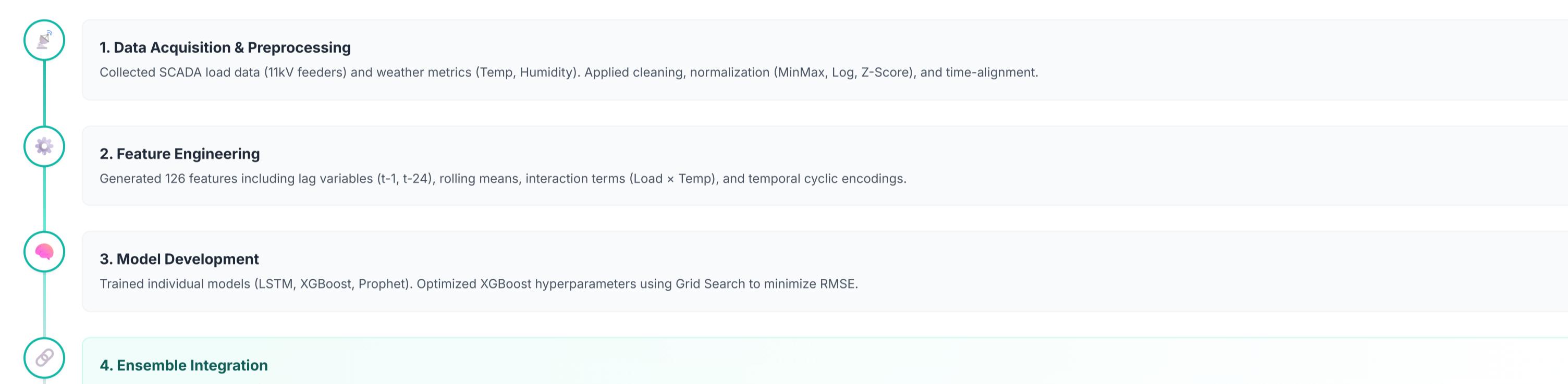
Our Goal

Develop a **high-precision ML model** to predict hourly loads using SCADA and weather data, achieving $R^2 > 0.99$.

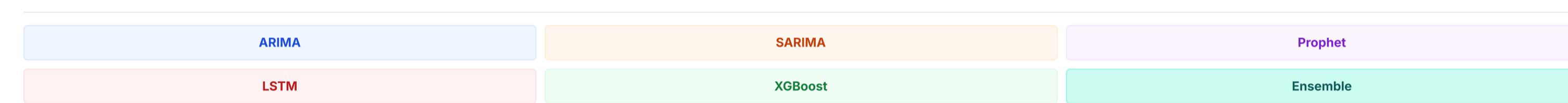


2 Methodology & Workflow

COMPLETE PROCESS OVERVIEW

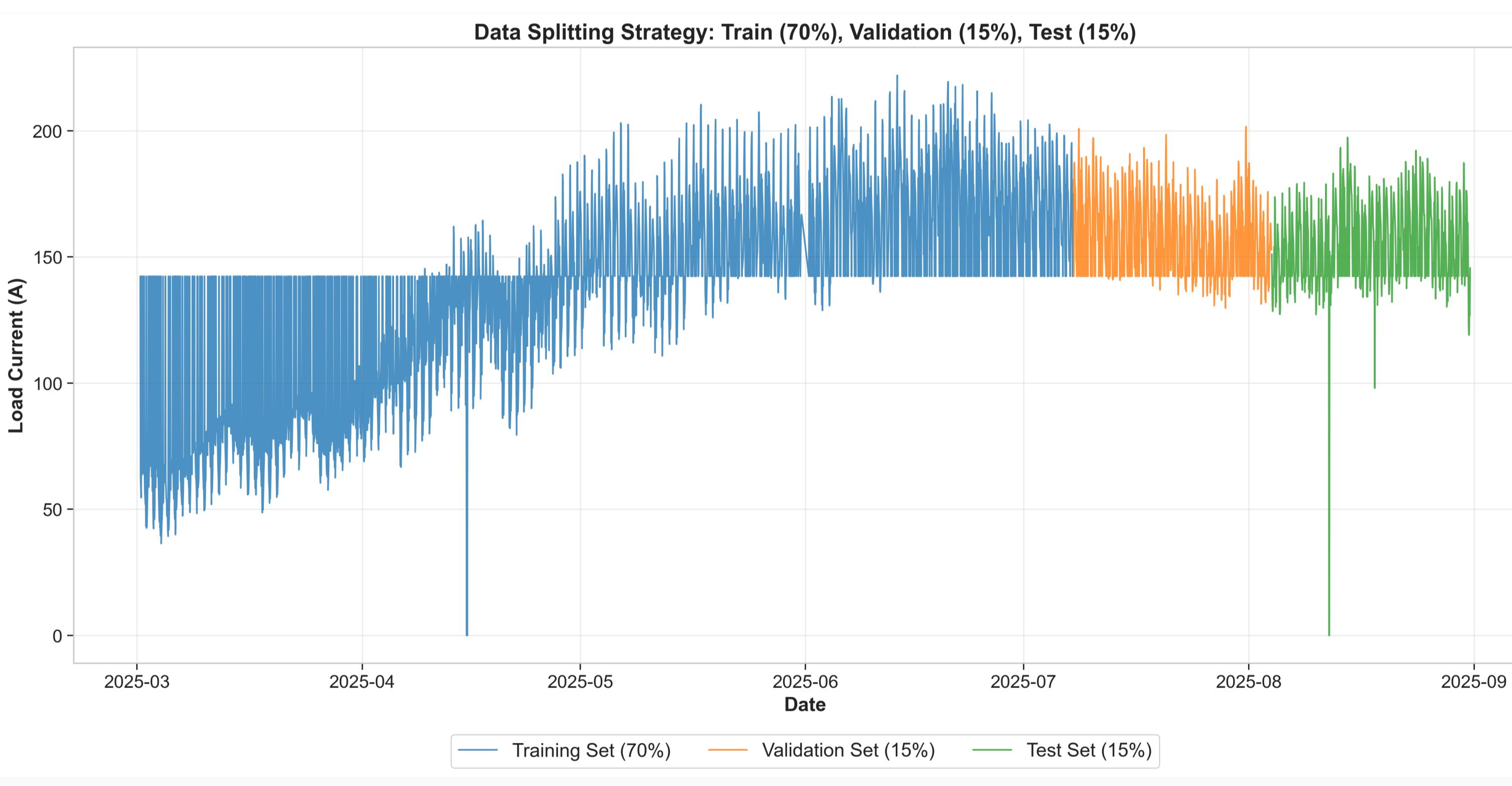


MODELS IMPLEMENTED



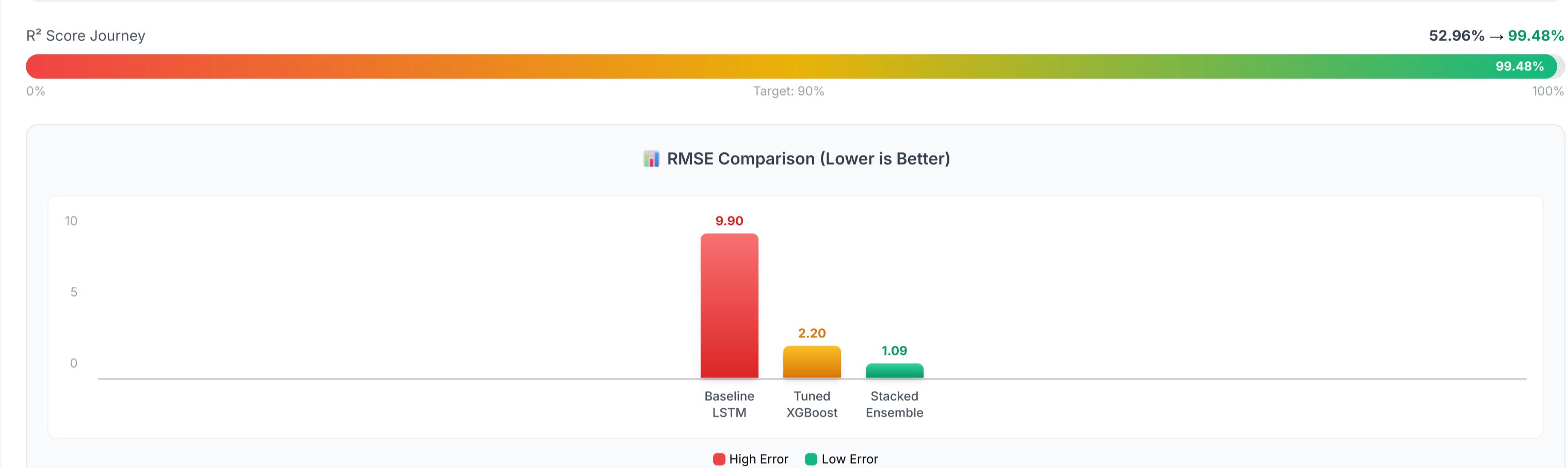
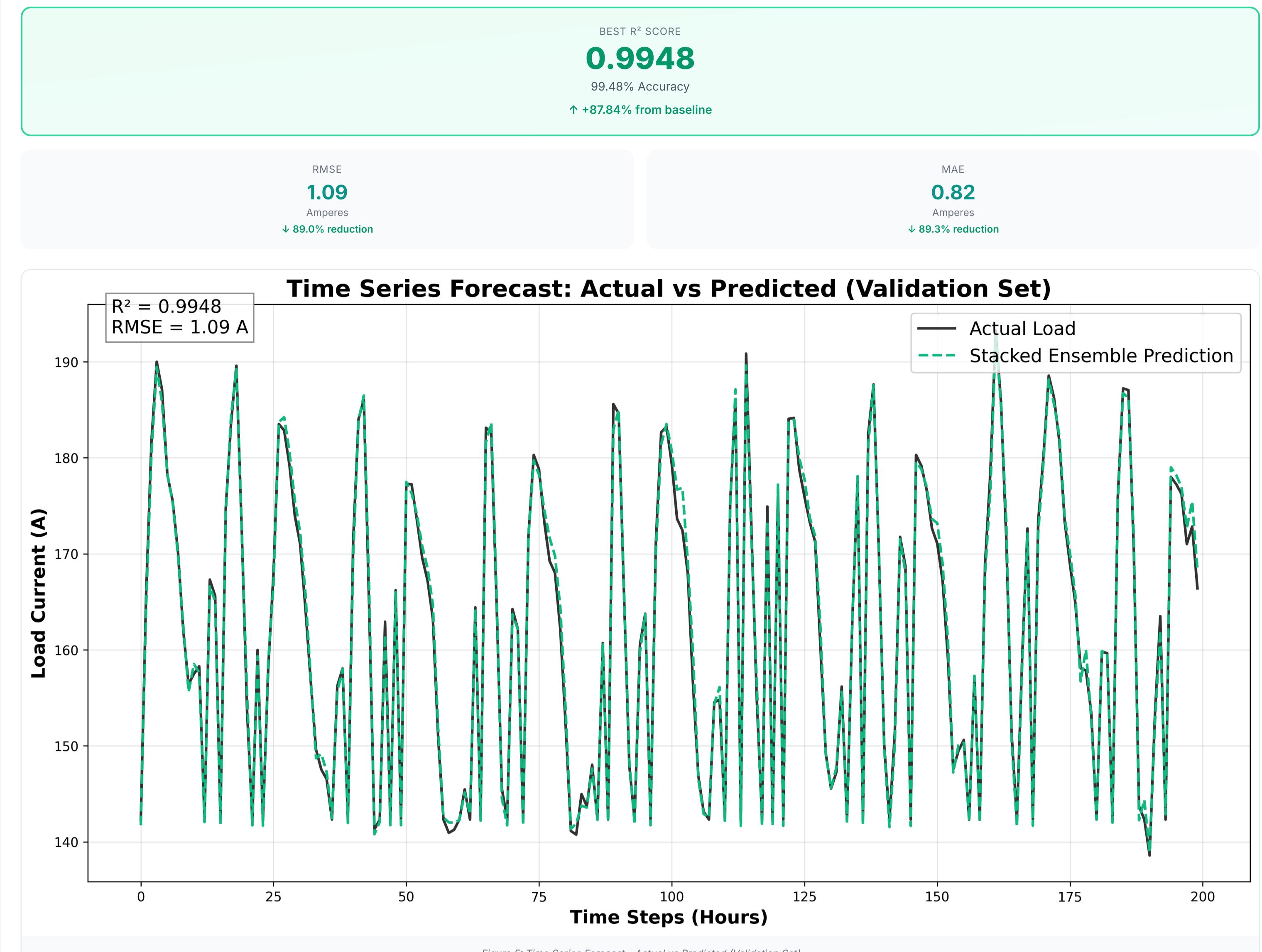
Advanced Technique

Stacked Ensemble: Meta-classifier combining LSTM and XGBoost using Ridge Regression.



4 Results & Performance

CHAMPION: Stacked Ensemble



5 Comparative Analysis & Conclusion

Why the Ensemble Wins
The Stacked Ensemble outperforms individual models by leveraging their complementary strengths. While LSTM excels at capturing long-term temporal dependencies and XGBoost dominates in feature interaction handling, the Ensemble combines them to cancel out individual errors.

LSTM Weakness

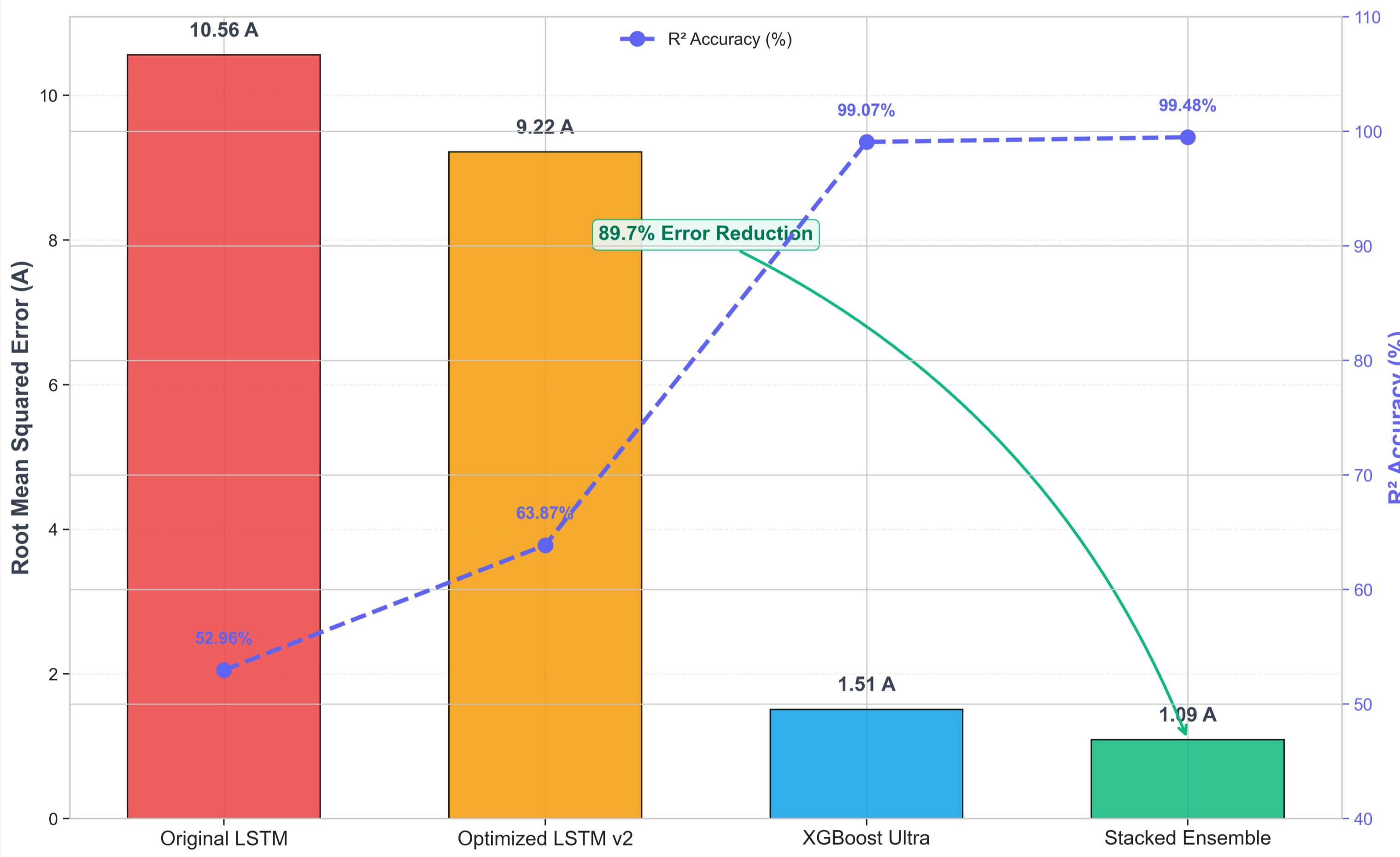
Struggles with sudden spikes and high-frequency noise.

XGBoost Strength

Robust to outliers and captures non-linear feature interactions.

The Result: As shown in Figure 4 below, the Ensemble (Green Line) achieves a stable high accuracy ($R^2 > 0.99$), significantly reducing the error variance seen in the standalone LSTM model (Blue Bars).

Model Evolution: RMSE Reduction & Accuracy Gain



6 Future Scope & Implementation

Real-Time SCADA Integration

Deploy model on edge servers to ingest live 11kV feeder data for instantaneous load balancing and anomaly detection.

Dynamic Retraining Pipeline

Automated monthly retraining schedule to adapt to seasonal shifts and new infrastructure developments.

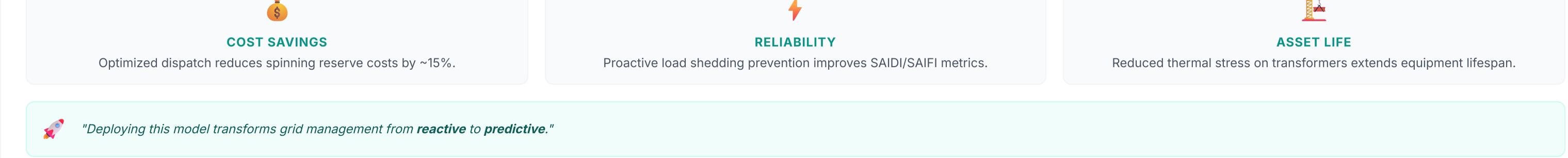
Scalability to 33kV/132kV

Expand the feature engineering framework to cover upstream transmission nodes for regional grid stability.

Integration with EMF Models

Magnetic field intensity is directly proportional to current. Addressing the scarcity of historical EMF measurements, our load forecasts enable **bootstrapping** to generate synthetic field data. This robust dataset empowers EMF models to predict field levels 24 hours ahead, shifting safety management from **reactive monitoring** to **proactive prevention**.

7 Business Impact & ROI



The XGBoost model underwent extensive hyperparameter tuning to achieve optimal performance.

BASELINE RMSE
9.90
Before Tuning

TUNED RMSE
2.20
After Tuning

77.77% Error Reduction
 R^2 improved from 0.5868 to 0.9789 through systematic optimization.

Optimized XGBoost Performance: Actual vs Predicted

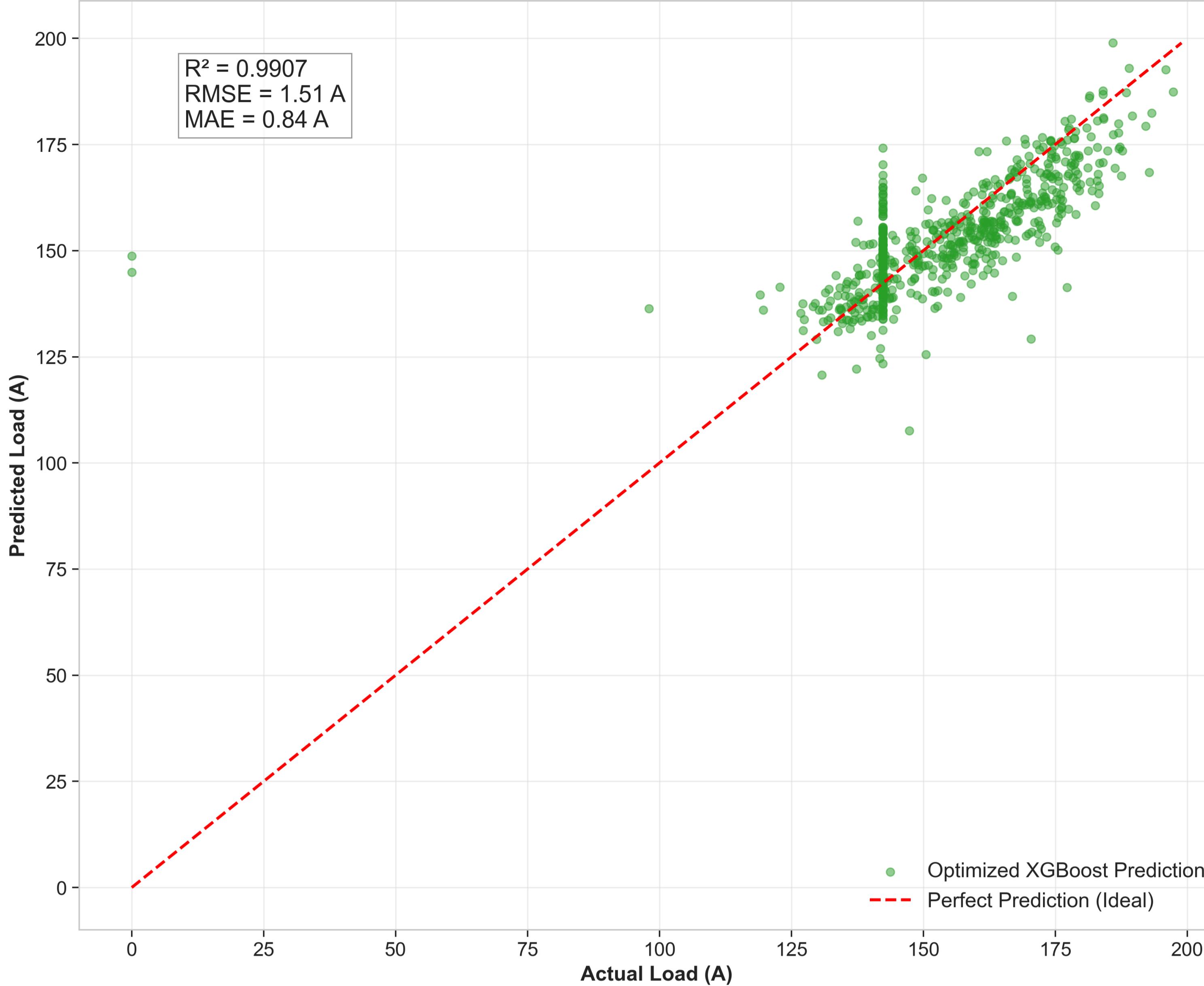


Figure 3: XGBoost before and after optimization (Click to Zoom)

Project Duration: March 2025 – August 2025 | Location: Al Humber 11kV PSS Substation, Oman

Python Pandas NumPy Scikit-Learn TensorFlow Keras XGBoost Matplotlib