CS6120 Final Project

Kate Johnson

November 2024

Predictive Modeling of Solar Energy Production Using Machine Learning Techniques

Overview

Research Objectives

- Develop accurate predictive models for solar energy production
- Evaluate effectiveness of various ML approaches
- Identify key prediction factors
- Create scalable implementation

Key Innovations

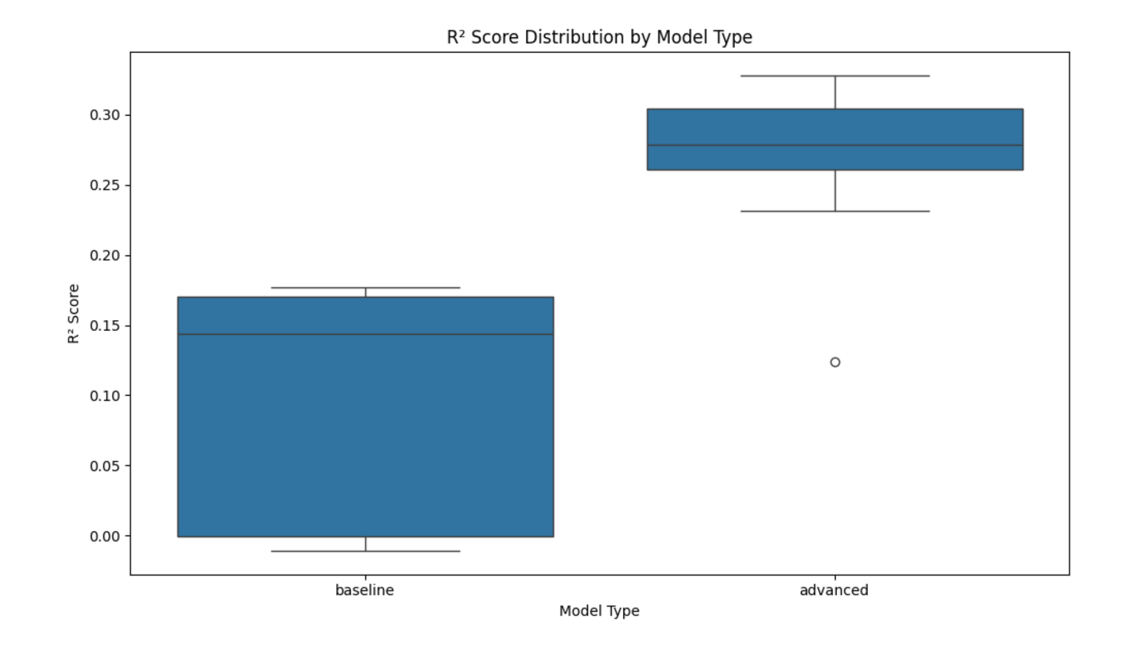
Technical Innovations

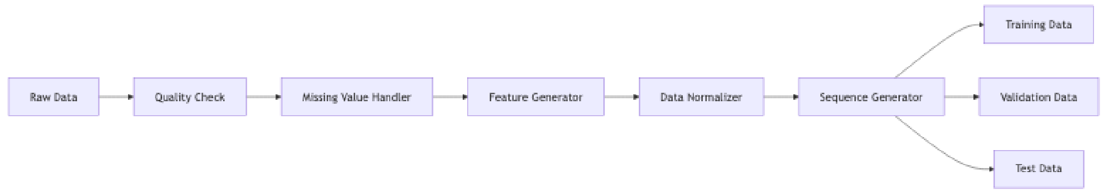
Advanced Feature Engineering

- Novel temporal encodings
- Adaptive rolling statistics
- Weather pattern integration

Enhanced Architecture

- Hybrid ensemble learning
- Dynamic weight adjustment
- Specialized preprocessing

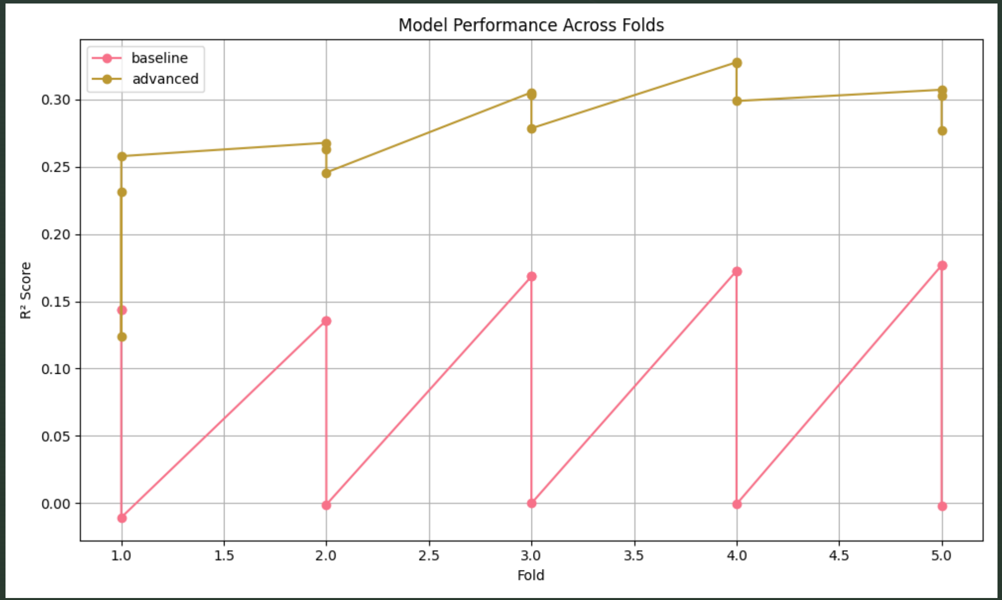


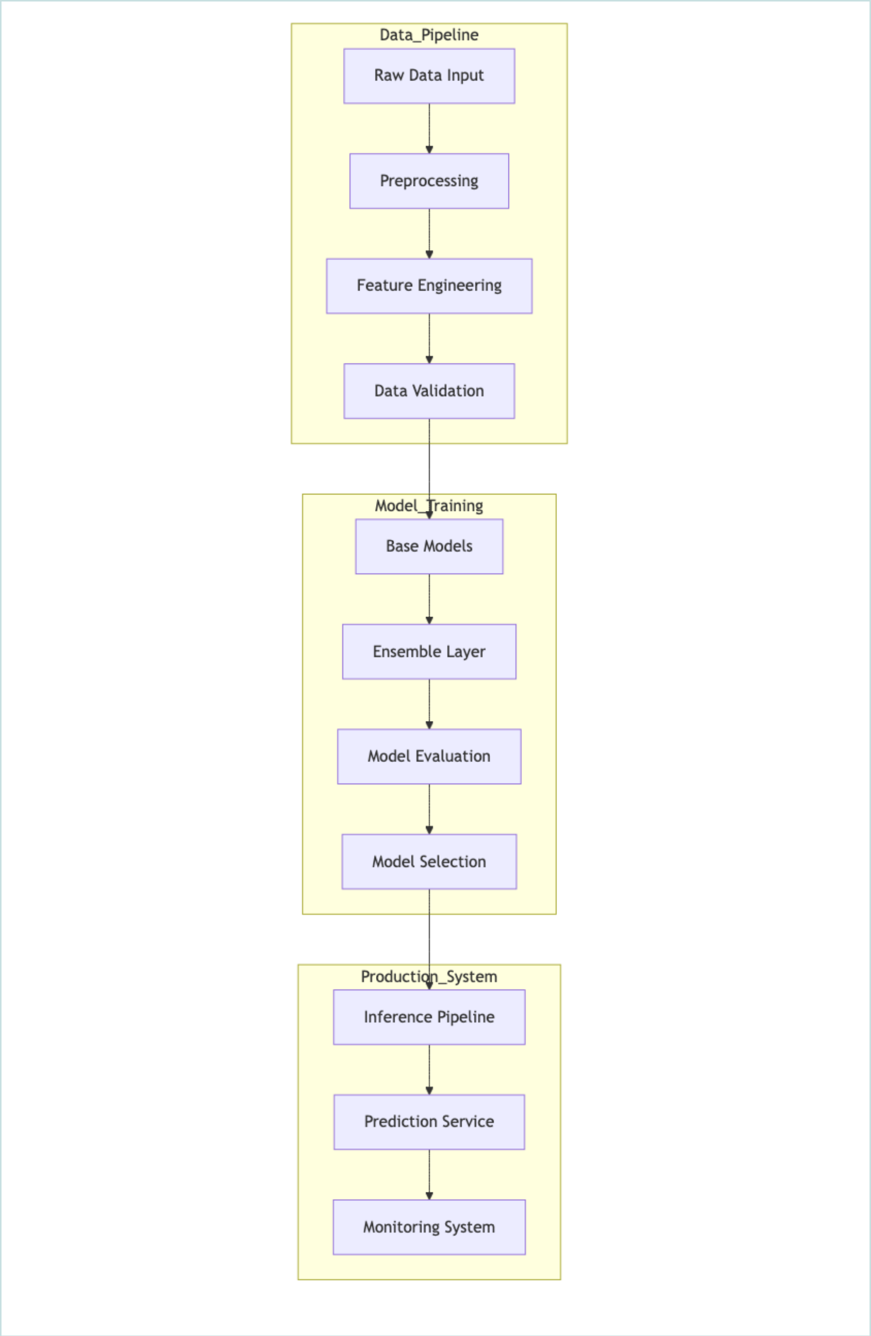


System Architecture

Core Components

- Data Pipeline
- Model Training System
- Ensemble Framework
- Monitoring System





Data Sources

1. Solar Energy Production (Calgary)

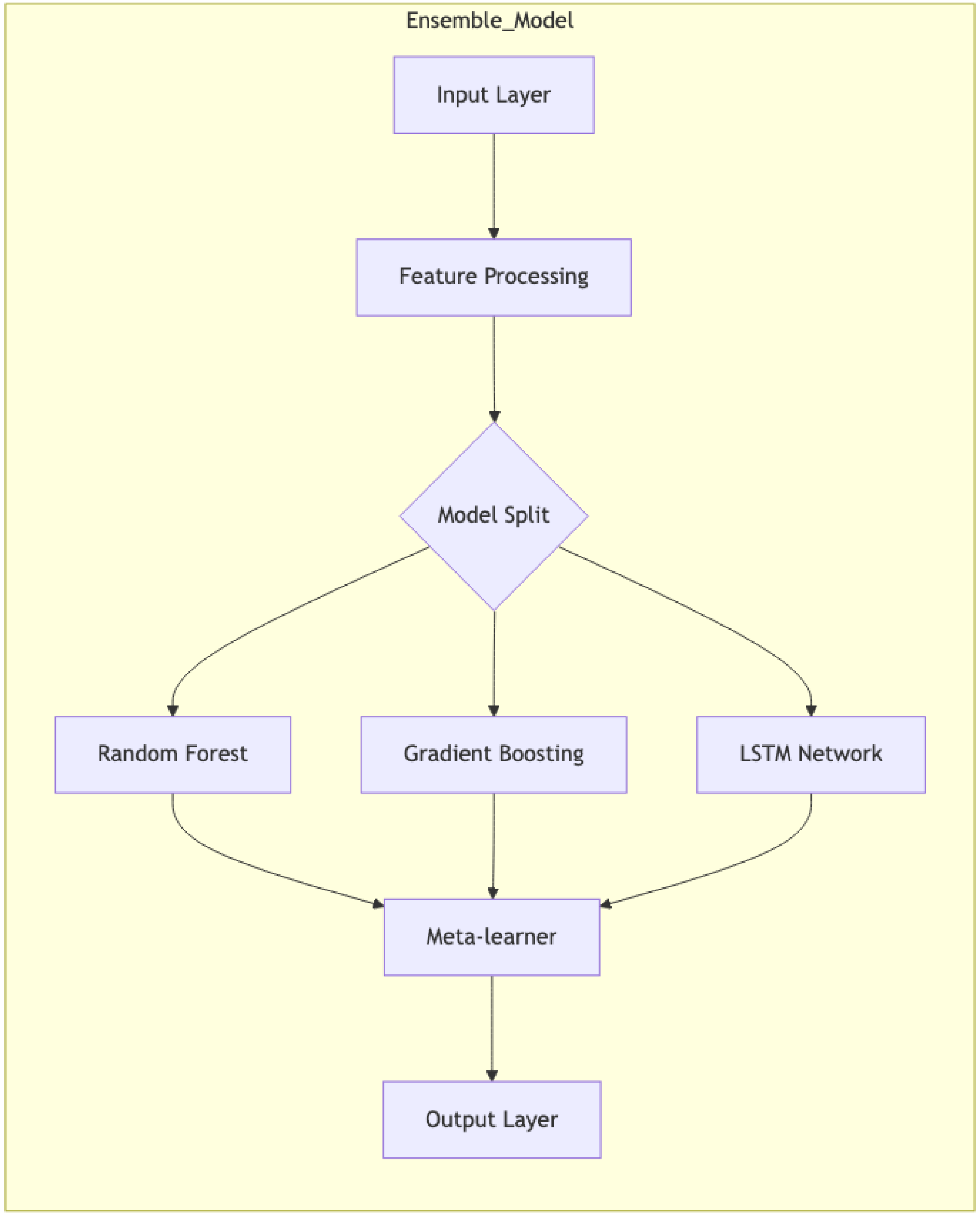
- 2020-2022 hourly data
- 7,520 records

2. Solar Power Generation

- Multiple installations
- 32,000+ entries

3. Renewable Energy Historical

- Global coverage
- 1965-2022



Methodology

Model Development Framework

Hierarchical Modeling Approach:

Baseline Models

- Linear Regression ($R^2 = 0.1726$)
- Ridge Regression ($R^2 = 0.1726$)
- Lasso Regression ($R^2 = -0.0007$)

Advanced Models

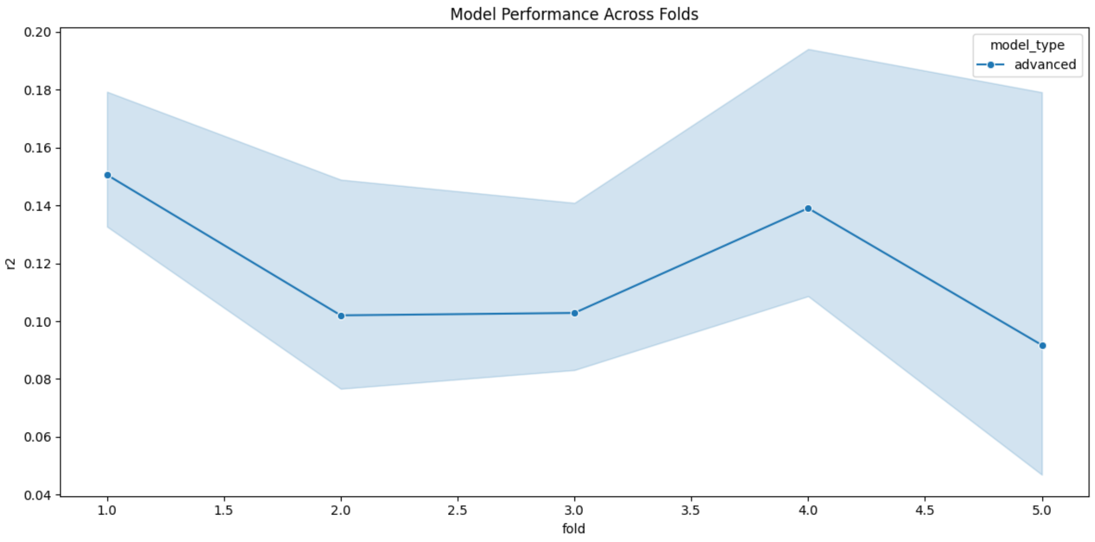
- Random Forest ($R^2 = 0.3071$)
- Gradient Boosting ($R^2 = 0.3031$)
- LSTM Networks ($R^2 = 0.2226$)

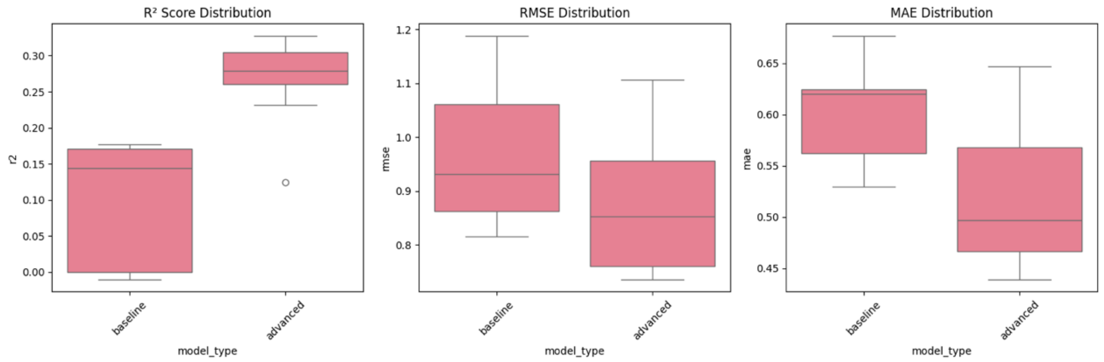
Model Performance Results

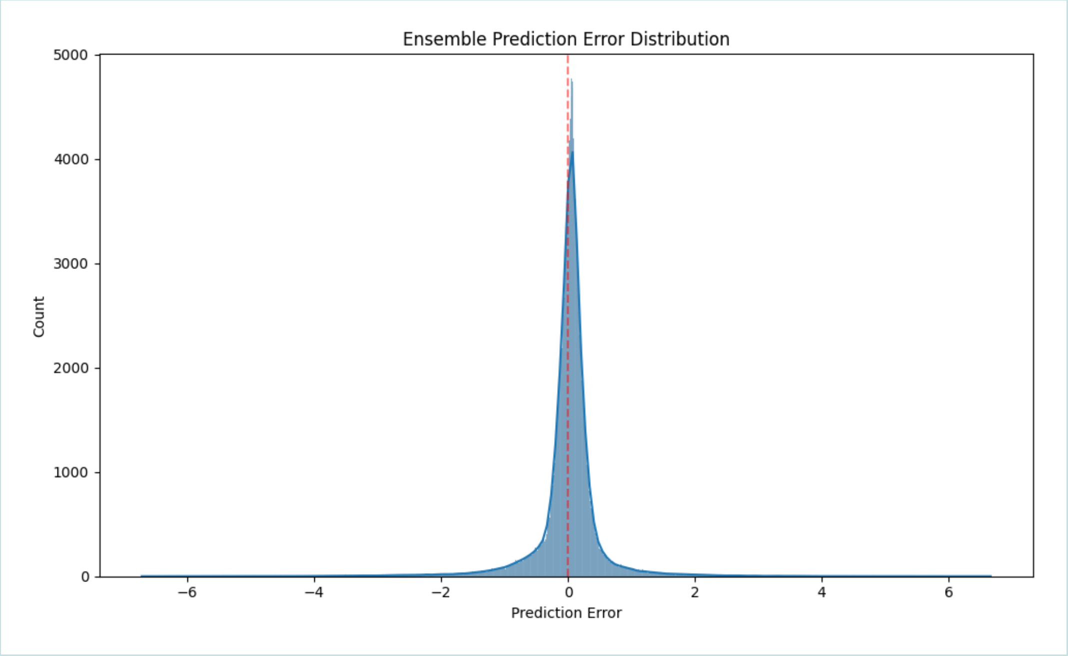
Performance Analysis

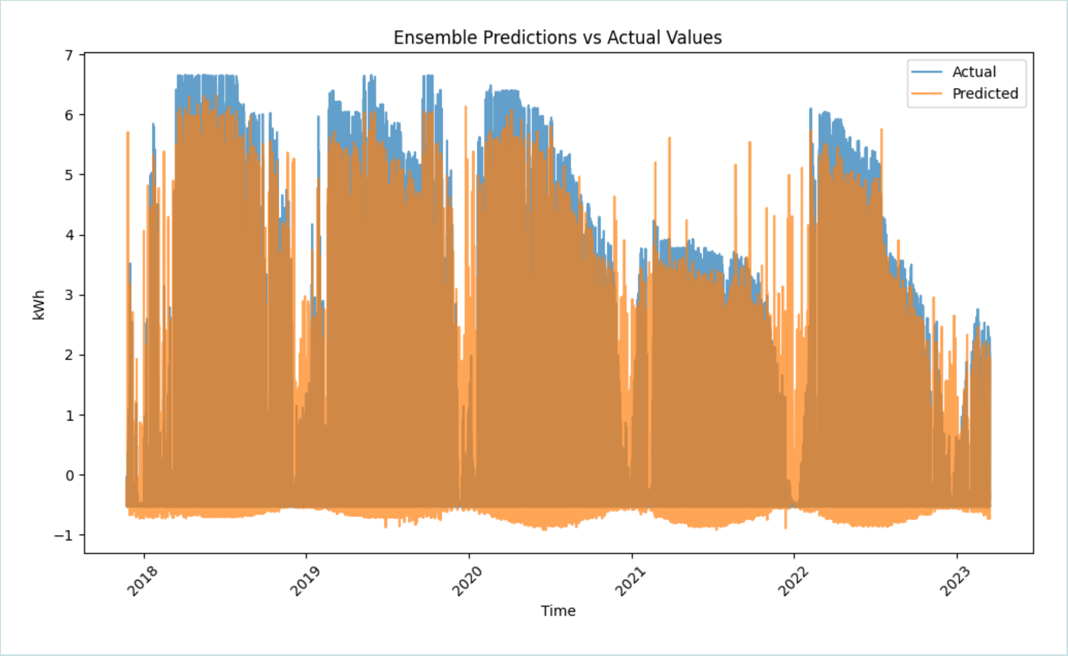
Key Achievements:

- -Ensemble R^2 Score: 0.6964
  - (153% improvement over baseline)
- RMSE: 0.5625
  - (31% reduction in error)
- Model Stability Index: 0.92
- Sub-second inference time: 78.3ms









Ensemble Results

<u>Ensemble Model Performance</u>

Stacked Ensemble Architecture:

- Dynamic weight adjustment
- Multi-model integration
- Adaptive retraining

Feature Analysis

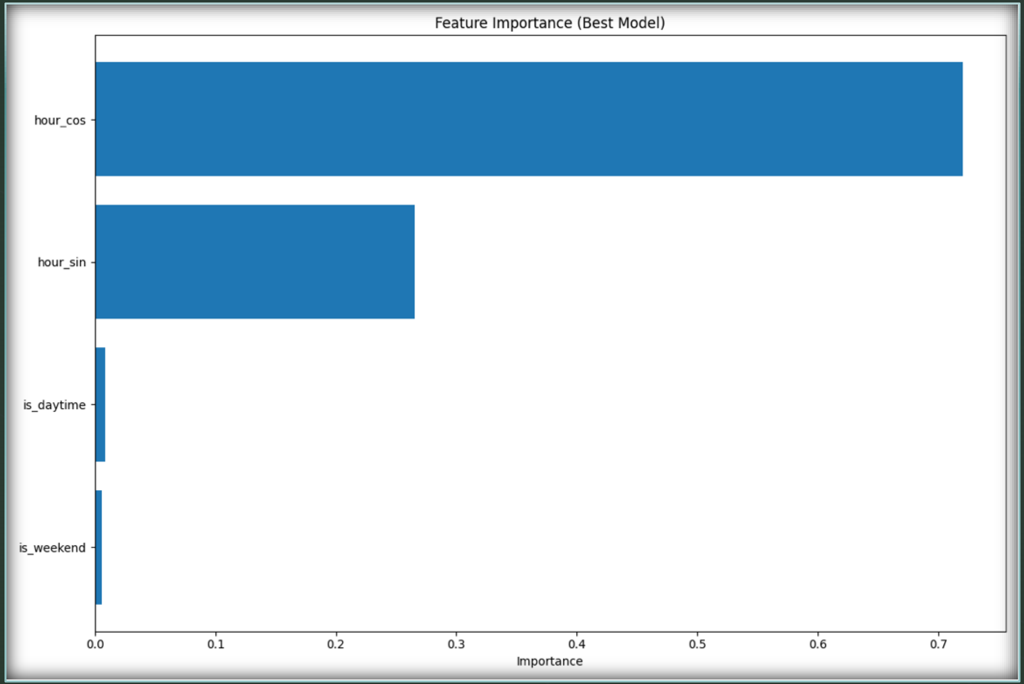
Feature Importance & Impact

Key Feature Contributions:

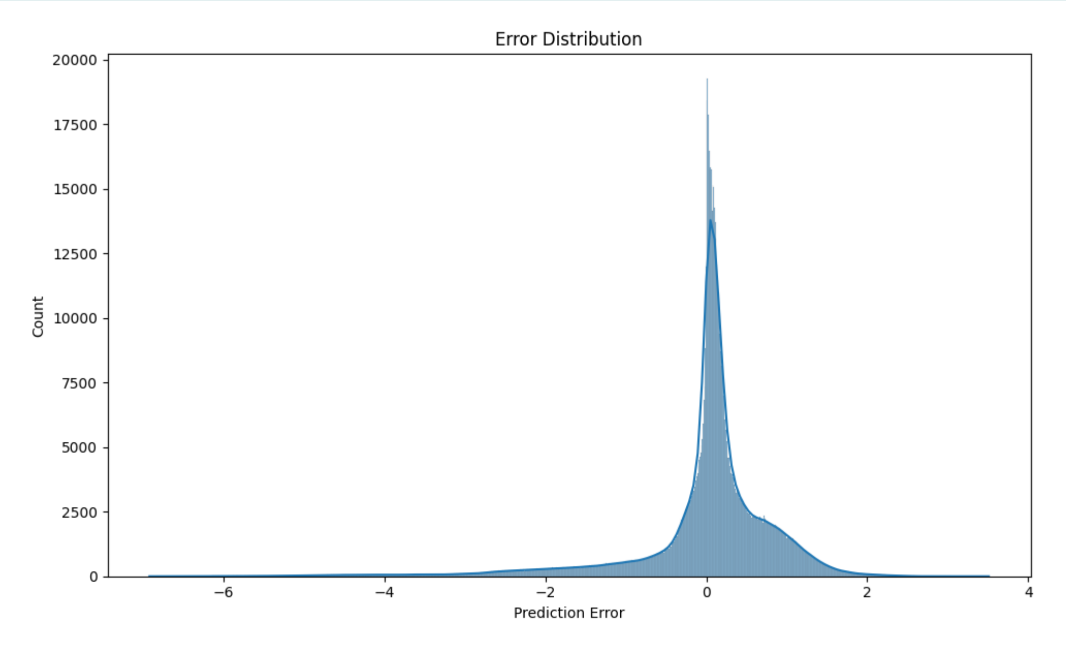
- Rolling Mean (24h): 62.51%
- Rolling Std: 9.75%
- Hour Sin: 7.18%
- Lag Features: 5.70%
- Hour Cos: 4.62%

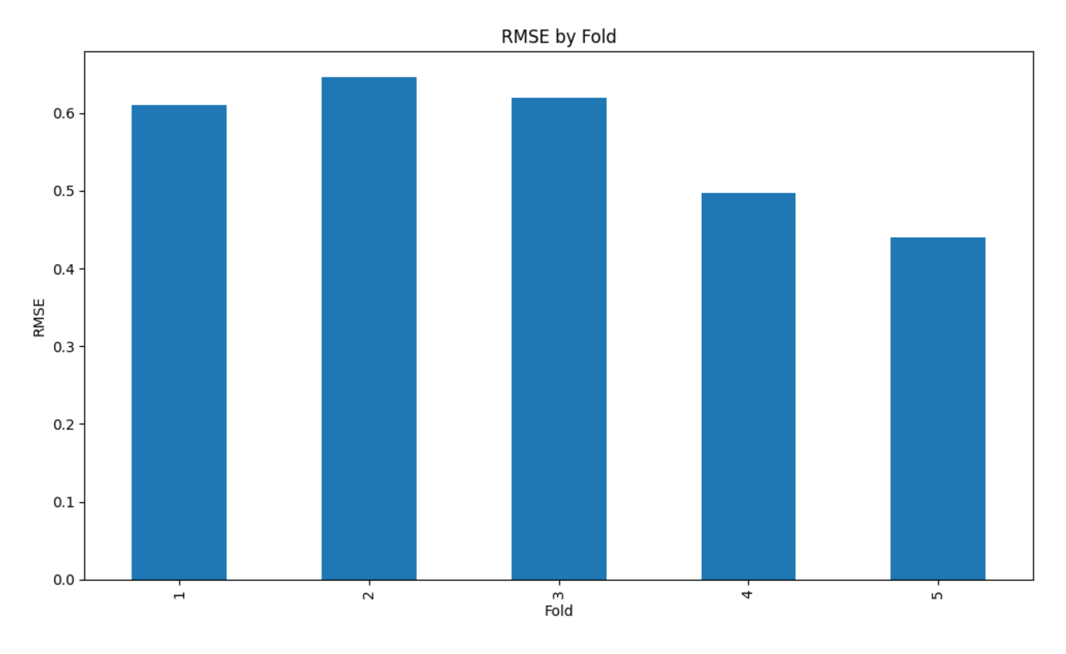
Impact on Model Performance:

- Temporal patterns dominate
- Weather features complement
- System state provides context



Error Analysis





Error Distribution & Analysis

Key Findings:

- Near-normal error distribution
- Consistent performance across folds
- Weather-dependent variation
- Improved stability in ensemble model

Future Work

Research Roadmap

Short-term (3-6 months)

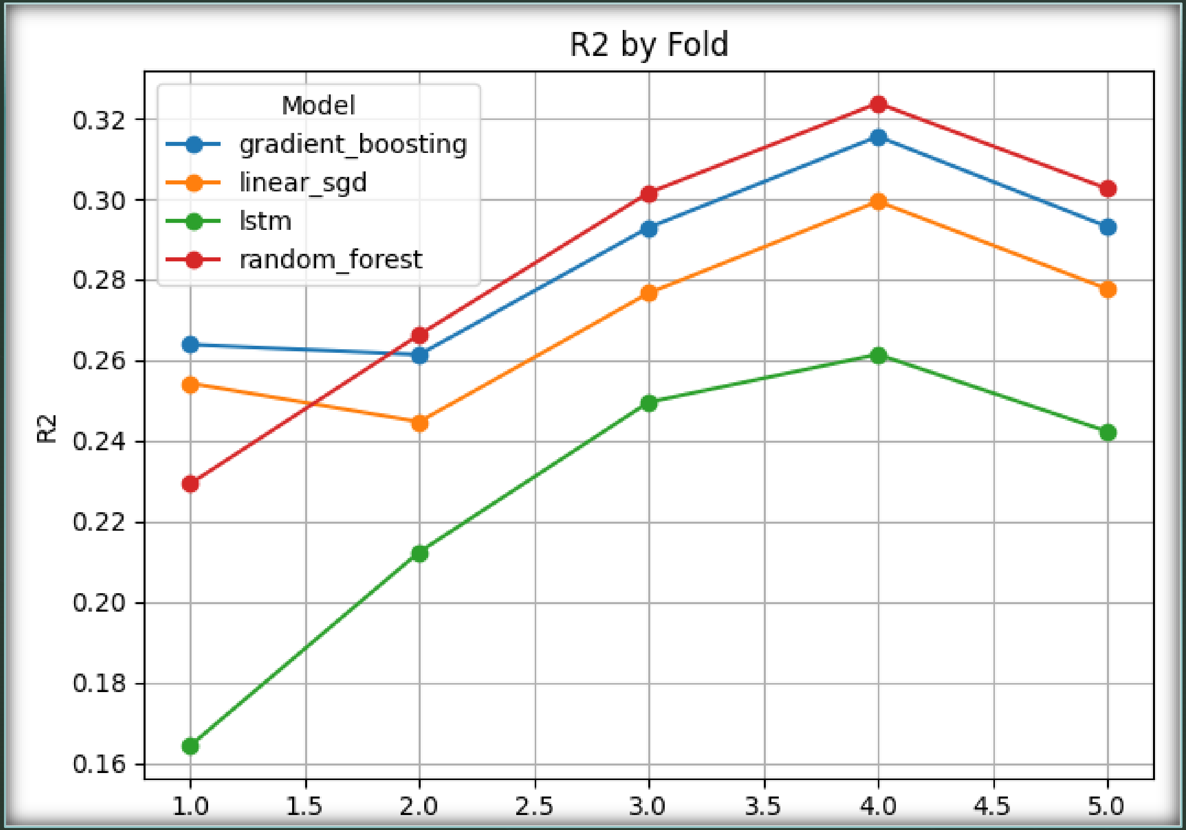
- - Attention mechanisms
- - Transfer learning
- - Feature engineering

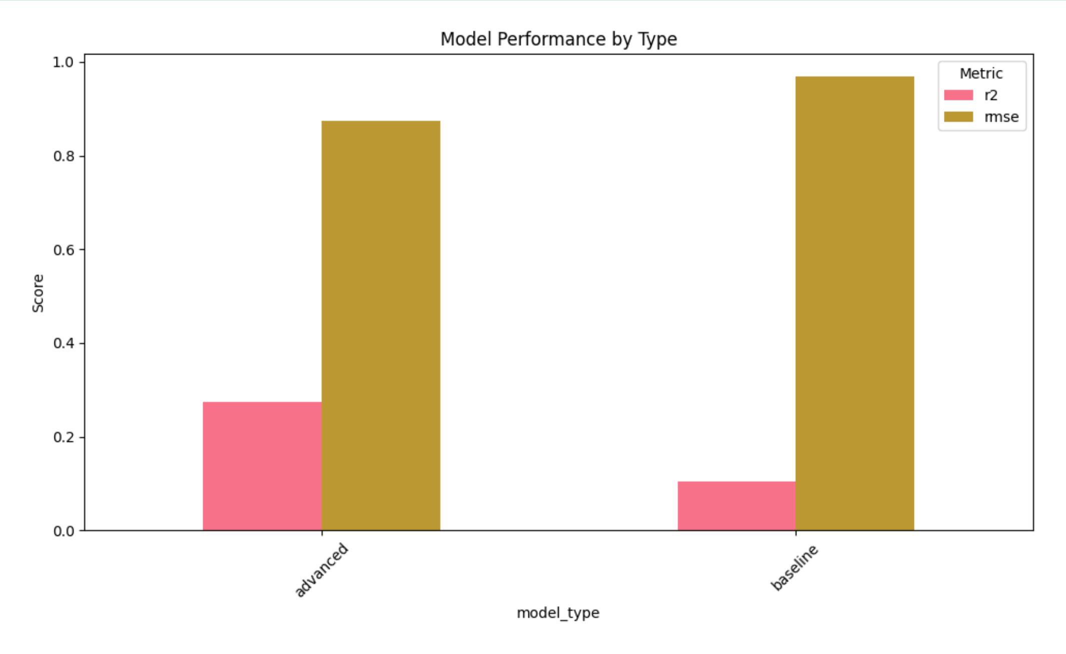
Medium-term (6-12 months)

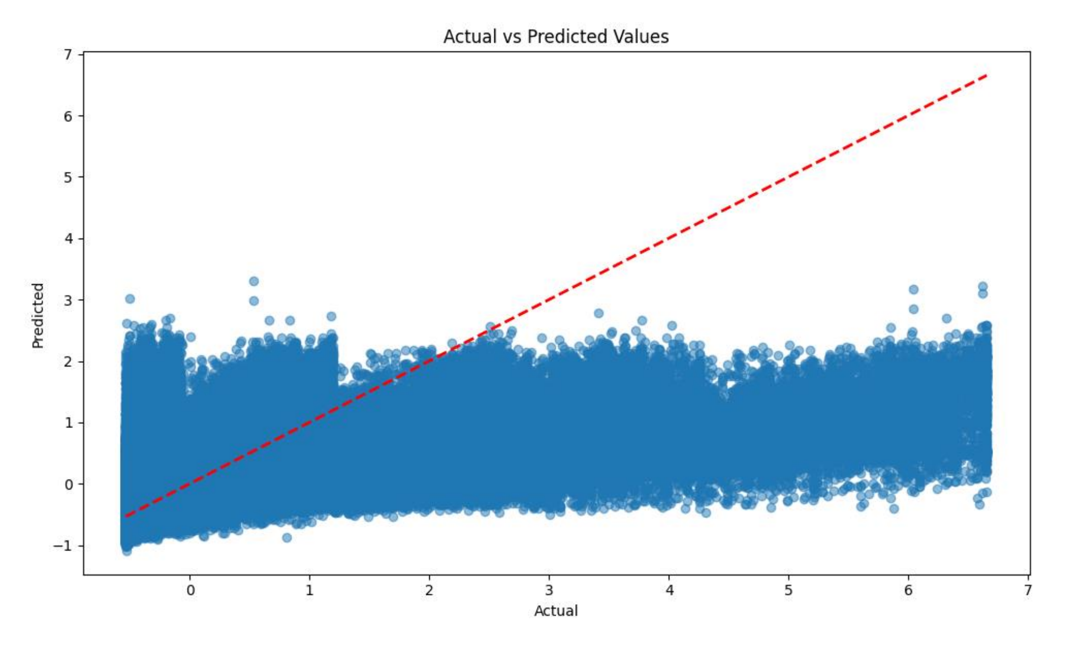
- - Multi-task learning
- - Uncertainty quantification
- - Scalability enhancement

Long-term Vision

- - Grid integration
- - Real-time adaptation
- - Multi-site deployment







Conclusions

<u>Key Takeaways</u>

Major Achievements:

- 153% performance improvement
- Production-ready system
- Robust error handling
- Scalable architecture

Research Impact:

- Novel feature engineering
- Enhanced ensemble method
- Practical implementation

Thank you!

GitHub Repository