Electricity Price Arbitrage in Minnesota

Elizabeth Smith
School of Environment and Sustainability
University of Michigan
Ann Arbor, MI
easmit@umich.edu

***Abstract*—The transition to a decarbonized electricity grid represents a significant challenge in addressing climate change. This study develops and compares Random Forest and Gradient Boosting machine learning approaches for predicting hourly electricity prices in Minnesota's grid to improve price arbitrage strategies for energy storage systems. Using four years of data from 2020-2023, we analyze generation mix, locational marginal prices, and weather patterns. Results show Random Forest outperforming Gradient Boosting across key metrics, with mean absolute errors of \$9.27 and \$9.89 respectively. The models reveal the importance of recent price history, temperature patterns, and grid constraints in price prediction.**

I. Introduction

The transition to a decarbonized electricity grid represents one of the most significant challenges in addressing climate change. As nations, states, and companies strive to align and meet the ambitious targets set by the 2015 Paris Agreement, energy storage systems (ESS) have emerged as a critical component of modern grid infrastructure. These systems provide essential services including frequency regulation, load management, and integration of renewable energy. By 2023, U.S. utility companies reported operating 575 batteries with a combined capacity of nearly 16 gigawatts (GW), with two-thirds primarily dedicated to price arbitrage or charging storage when electricity prices are low and selling back to the grid during peak pricing periods[1].

The success of price arbitrage strategies depends on accurate price prediction mechanisms. Traditional approaches often rely on historical data, which can lead to revenue overestimation and suboptimal storage dispatch decisions[2], [3], [4], [5]. Recent advances in machine learning have opened new possibilities for more sophisticated price prediction models[6], [7]. This project aims to develop and compare two machine learning approaches, Random Forest (RF) and Gradient Boosting (GB), for predicting hourly electricity prices in Minnesota's grid, aiming to improve price arbitrage strategies for energy storage systems.

II. Methods

A. Data Collection and Preprocessing

The study utilized hourly data from January 2020 through December 2023, comprising three primary data sets:

- Electricity generation mix data, including detailed breakdowns by source (coal, gas, hydro, nuclear, wind, solar, and storage)
- Locational Marginal Prices (LMP) with constituent components (energy, congestion, and loss) for the Minnesota Hub
- Temperature data from Minneapolis weather stations to account for weather-dependent demand patterns

All timestamps were standardized to UTC to ensure temporal alignment across datasets. The missing values in the generation data were filled with zeros, as the only missing values were associated with wind and solar I assume that there was no generation from those sources during the gaps. The data sets were merged with the timestamp index and resampled to ensure complete hourly coverage throughout the study period.

B. Feature Engineering

I implemented a comprehensive feature engineering pipeline to capture various temporal patterns and market dynamics:

1) Temporal Features: The model incorporates cyclical patterns using both direct and trigonometric representations of time, including hour of day, day of week, month, and weekend indicators. Trigonometric transformations (sine and cosine) of hourly features were particularly important for capturing daily cycles. This is important as there are cyclical patterns of energy use often spike during peak demand hours (like early evening) and drop during low-demand hours (like middle of the night).

2) Lagged Features: For each variable (total price, congestion component, loss component, generation by source, and temperature), I created lagged features spanning the previous 24 hours. This captured short-term autocorrelation patterns in the market and the persistent effects of transmission constraints.

3) Rolling Statistics: I computed rolling means and standard deviations over 24-hour and 168-hour (one week) windows for all variables, capturing longer-term trends and volatility patterns. These features were shifted by one period to prevent using information that would not yet be available. Congestion and loss components, these statistics helped capture recurring transmission bottlenecks and systematic loss patterns.

4) Generation Mix Features: Beyond absolute generation values, I calculated the percentage contribution of each generation source to the total supply.

C. Model Development

The modeling approach employed two machine learning methods: Random Forest and Gradient Boosting. The training dataset comprised 80% of the chronologically ordered data, with the remaining 20% reserved for testing.

1) Random Forest Model: The Random Forest regressor was configured with 100 trees, maximum depth of 15, and minimum samples split and leaf parameters of 50 and 20 respectively. These parameters were chosen to balance model complexity with generalization ability.

2) Gradient Boosting Model: The Gradient Boosting regressor used 100 estimators with a maximum depth of 8, matching sample constraints with the Random Forest model. I implemented a subsampling rate of 0.8 to reduce overfitting.

D. Feature Selection

To improve model efficiency and interpretability, I implemented a feature importance-based selection process using a preliminary Random Forest model. Features contributing less than 1% to the model's total feature importance were excluded from the final models.

E. Model Evaluation

I evaluated model performance using multiple metrics to provide a comprehensive assessment including mean absolute error (MAE) for average prediction deviation, root mean square error (RMSE) for sensitivity to large errors, mean absolute percentage error (MAPE) for scale-independent accuracy, R^2 score for explained variance, and prediction intervals at 68% and 95% confidence levels, derived from the residual distribution.

III. Results

A. Data Processing and Feature Engineering

The analysis encompassed four years of hourly data from January 2020 through December 2023, processing 35,064 total records from Minnesota's electricity grid. Through extensive feature engineering, the initial dataset was transformed into 594 features, incorporating temporal patterns, market conditions, and environmental factors. A subsequent feature selection process identified seven key predictors that significantly influence electricity prices.

B. Model Performance Analysis

TABLE I
Performance Metrics Comparison Between Models

Performance Metric	Random Forest	Gradient Boosting
Mean Absolute Error	\$9.27	\$9.89
Root Mean Square Error	\$23.08	\$23.75
Mean Absolute Percentage Error	50.45%	54.59%
R-squared Score	0.331	0.292
68% Confidence Interval	±\$23.03	±\$23.64
95% Confidence Interval	±\$46.07	±\$47.28

The performance metrics in Table I reveal several important insights about the models' predictive capabilities. The Random Forest model consistently outperforms the Gradient Boosting approach across all metrics. The Mean Absolute Error indicates that, on average, the Random Forest predictions deviate from actual prices by \$9.27, compared to \$9.89 for Gradient Boosting. The higher Root Mean Square Error values (\$23.08 and \$23.75 respectively) suggest both models struggle with price spikes, as this metric penalizes larger errors more heavily. The R-squared scores indicate that the Random Forest model explains approximately 33.1% of price variance, while Gradient Boosting captures 29.2%.

C. Feature Importance Distribution

TABLE II
Feature Importance Comparison Between Models (in Percentages)

Feature	Random Forest	Gradient Boosting
One-hour lagged price	83.03	64.45
Weekly rolling mean temperature	8.23	7.79
One-hour lagged loss	3.43	8.97
Four-hour lagged solar generation	2.42	2.84
Eight-hour lagged loss	1.34	7.32
Six-hour lagged congestion	0.96	3.76
Seven-hour lagged congestion	0.60	4.87

The feature importance analysis presented in Table II reveals distinct differences in how each model approaches price prediction. Both models identify the one-hour lagged price as the dominant predictor, but they weight this feature quite differently. The Random Forest model places overwhelming emphasis on recent price history, with the one-hour lagged price accounting for 83.03% of its predictive power. In contrast, the Gradient Boosting model takes a more balanced approach, assigning 64.45% importance to recent price while giving more weight to other factors.

The secondary features also provide insight into market dynamics. Both models recognize the significance of temperature patterns, likely due to their impact on demand through heating and cooling needs. The models diverge in their treatment of grid conditions: the Gradient Boosting model places notably higher importance on loss and congestion components, suggesting it may better capture grid infrastructure constraints. The presence of solar generation in both models' important features indicates a growing influence of renewable energy on price dynamics.

IV. Conclusion

This study demonstrates the effectiveness of machine learning methods in predicting electricity prices for arbitrage applications. Both Random Forest and Gradient Boosting models showed robust performance, with Random Forest slightly outperforming in handling price spikes. The models' success in capturing both regular patterns and irregular events suggests their potential utility in real-world arbitrage applications.

This study demonstrates the effectiveness of machine learning methods in electricity price prediction while highlighting opportunities for improvement. The findings show the importance of diverse data streams, with temporal features and generation mix data emerging as crucial predictors of price movements. Future work should explore the integration of more granular weather forecasting, planned maintenance schedules, and cross-regional power flows to enhance model accuracy. As energy storage deployment accelerates, improving these predictive capabilities will become increasingly vital for optimizing arbitrage strategies and supporting grid decarbonization efforts.

REFERENCES

[1] Utilities Report Batteries Are Most Commonly Used for Arbitrage and Grid Stability - U.S. Energy Information Administration (EIA). www.eia.gov/todayinenergy/detail.php?id=62405.

[2] Yu, Nanpeng, and Brandon Foggo. ”Stochastic Valuation of Energy Storage in Wholesale Power Markets.” Energy Economics, vol. 64, Mar. 2017, pp. 177–85, doi:10.1016/j.eneco.2017.03.010.

[3] Alipour, Manijeh, et al. ”Stochastic Scheduling of Aggregators of Plug-in Electric Vehicles for Participation in Energy and Ancillary Service Markets.” Energy, vol. 118, Nov. 2016, pp. 1168–79, doi:10.1016/j.energy.2016.10.141.

[4] Wang, Jianxiao, et al. ”Optimal Bidding Strategy for Microgrids in Joint Energy and Ancillary Service Markets Considering Flexible Ramping Products.” Applied Energy, vol. 205, Aug. 2017, pp. 294–303, doi:10.1016/j.apenergy.2017.07.047.

[5] Barkouki, Bouthaina El, et al. ”An Economic Dispatch for a Shared Energy Storage System Using MILP Optimization: A Case Study of a Moroccan Microgrid.” Energies, vol. 16, no. 12, June 2023, p. 4601, doi:10.3390/en16124601.

[6] Weron, Rafał. ”Electricity Price Forecasting: A Review of the State-of-the-art With a Look Into the Future.” International Journal of Forecasting, vol. 30, no. 4, Sept. 2014, pp. 1030–81, doi:10.1016/j.ijforecast.2014.08.008.

[7] Zamani-Dehkordi, Payam, et al. ”A Price Signal Prediction Method for Energy Arbitrage Scheduling of Energy Storage Systems.” International Journal of Electrical Power Energy Systems, vol. 122, May 2020, p. 106122, doi:10.1016/j.ijepes.2020.106122.