Detecting Single Point Attacks in Electricity Markets using Decision Trees

Ioannis Vourkas

Abstract

Electricity market operators face increasing risks from both unintentional errors and deliberate attacks that can manipulate dispatch decisions and inflate costs. This paper develops a machine learning framework to detect single-point market anomalies using synthetic data generated from the IEEE 300-bus test case. We simulate four types of attacks targeting generator parameters: ramp rate manipulation, upper/lower generation limit alterations, and cost coefficient modifications. Through comprehensive feature engineering focusing on temporal dynamics of optimal power flow solutions, we train and evaluate six classification algorithms. Our best-performing model, a Voting Classifier ensemble, achieves 92.45% accuracy on the full dataset. Seasonal analysis reveals interesting patterns: while Gradient Boosting consistently outperforms other models on limited seasonal data (81.25%-89.58% accuracy), ensemble methods excel with larger, diverse datasets. The most informative features capture changes in objective function values (`fval_change`, `fval_diff_lag_1`), highlighting the economic signatures of market manipulations. Our findings demonstrate that machine learning can effectively identify market anomalies and provide a foundation for real-time monitoring systems to enhance electricity market integrity and protect consumers from financial harm.

Keywords: **electricity markets**, **market anomaly detection**, **machine learning**, **optimal power flow**, **cyber-physical security**, **market integrity**, **gradient boosting**, **ensemble methods**

1 Introduction

Power system operators face increasingly complex challenges in maintaining grid reliability and market integrity in today's evolving energy landscape. As renewable energy sources become more prevalent and demand response programs grow in sophistication, system operators must adapt their forecasting and modeling techniques to ensure accurate market operation. However, these adaptations sometimes lead to unforeseen consequences with significant financial implications.

A stark example of such consequences occurred in Ontario, Canada, where the Independent Electricity System Operator (IESO) inadvertently introduced "phantom demand" into their electricity market from May 2016 to April 2017 [1]. During this period, as reported by the Ontario Energy Board, the system operator erroneously counted up to 220 MW of fictitious demand - equivalent to the power requirements of a small city. This error arose when IESO began integrating small-scale embedded generation (primarily solar) into its wholesale market model and implementing demand response programs. The miscalculation assumed maximum consumption at sites without meters and double-counted that consumption, resulting in artificially inflated market prices.

The impact was substantial: the phantom demand "regularly inflated" the Hourly Ontario Energy Price (HOEP) by as much as $4.50/MWh on average, with dramatic price spikes during peak demand periods. In one instance, the hourly rate reached $1,619/MWh - the fourth highest in the history of Ontario's wholesale electricity market. The Ontario Energy Board estimated this error cost consumers between $450 million and $560 million in excess charges. Particularly affected were exporters and certain end-users who did not benefit from off-

setting reductions in global adjustment rates.

Perhaps most concerning was the IESO's failure to publicly disclose the error after its discovery. This lack of transparency, highlighted by the Market Surveillance Panel, undermines market confidence and prevents stakeholders from making informed decisions about energy consumption and investment.

To address such challenges and prevent similar incidents in the future, our research focuses on developing robust anomaly detection methods for electricity market operations. We created synthetic power system data using the IEEE 300-bus test case from the Power Grid Library (PGLib) [2]. Our simulation framework performs DC Optimal Power Flow (OPF) calculations under normal operating conditions and simulates various system attacks. This approach synthesizes data that captures both normal market behaviors and potential anomalies, providing a rich dataset for training machine learning models to detect market irregularities.

Our multi-model machine learning approach classified market conditions as normal or anomalous after implementing comprehensive feature engineering to capture temporal patterns and system characteristics. The best-performing model achieved 92.45% accuracy on the full dataset, while seasonal subset performance ranged from 81.25% to 89.58%. This performance variation underscores the critical role of seasonal patterns in power system behavior. Notably, while accuracy differed across seasonal subsets (models trained on season-specific data), Gradient Boosting consistently emerged as the top performer across all configurations.

This paper presents our methodology for synthetic data generation, feature engineering techniques, the entire pipeline we used for our training, model selection and evaluation results. We discuss the implications of our findings for market operators and regulators, and propose a framework for implementing anomaly detection systems in electricity markets to enhance transparency, fairness, and efficiency. The remainder of this paper is structured as follows: Section 2 provides a literature review of relevant work in power system anomaly detection, Section 3 details our methodology for data generation and model development, Section 4 presents our results and offers a discussion of the implications, and Section 6 concludes with recommendations for future research and implementation.

2 Literature Review

Anomaly detection in power systems has emerged as a critical area of research due to its implications for system reliability, economic stability, and security. A variety of recent approaches focus on identifying anomalies in electricity consumption or market price data using machine learning techniques.

One notable study proposes a two-stage deep learning method to detect anomalies in electricity consumption data one hour ahead of time, leveraging both Long Short-Term Memory (LSTM) forecasting and LSTM-based autoencoders to capture deviations from normal consumption behavior [3]. The study emphasizes the value of weather, temporal, and lag features in enhancing model performance, using Exponential Moving Average (EMA) thresholds to distinguish between local and global anomalies. While effective in identifying consumption-based anomalies, this approach is largely retrospective and focused on end-user demand patterns, rather than systemic market-level irregularities.

Another study investigates actual bidding behavior in deregulated electricity markets, addressing the limitations of conventional theoretical models that fail to reflect real-world dynamics [4]. The authors develop a data-driven framework that applies adaptive clustering (K-medoids) and Wasserstein distance metrics to extract meaningful bidding patterns from historical data. Although this work brings valuable insight into market participant behavior, it does not attempt real-time anomaly detection or offer predictive mechanisms to catch financially impactful market miscalculations, such as those caused by modeling errors.

A third study introduces a probabilistic anomaly detection framework targeting cyber-induced anomalies in electricity markets [5]. The method combines LSTM-based deterministic price forecasting with probabilistic modeling to identify deviations in locational marginal prices (LMPs), tested under simulated cyberattacks like load redistribution and price-responsive attacks. While this work makes significant contributions in linking cyber threats to market distortions, it is focused

on price anomalies under hypothetical attack scenarios rather than modeling or forecasting errors made by market operators in real-world cases.

In contrast to these existing approaches, our project is motivated by a real incident with significant financial repercussions: the introduction of phantom demand in the Ontario electricity market, which inflated prices and caused hundreds of millions in consumer losses due to market modeling errors. Unlike prior works that focus on either consumption-based anomalies, theoretical market behavior, or cyberattack detection, our approach targets anomalies that stem from operational or modeling flaws introduced by the system operator themselves—a largely overlooked but critically important domain.

We employ synthetic data generation using the IEEE 300-bus test case to simulate a wide variety of normal and anomalous market conditions, including those that mimic the kind of modeling flaws observed in Ontario. Our dataset is engineered with temporal features and tested across seasonal variations, and we evaluate multiple machine learning models to detect anomalies with high accuracy. This design not only enables the detection of anomalies in real-time market operations, but also directly addresses the risks posed by system-level modeling assumptions—something not tackled in prior literature.

By building models that can flag anomalies resulting from flawed integrations (e.g., embedded generation, metering gaps, or demand response miscalculations), we propose a proactive tool for operators and regulators to safeguard both market fairness and economic efficiency.

3 Methodology

3.1 Data Creation

To investigate the detection of anomalies in electricity markets, we developed a comprehensive simulation framework based on the IEEE 300-bus test case from the Power Grid Library (PGLib). Our approach centered on generating synthetic but realistic power system data that captured both normal operating conditions and various attack scenarios. This section details the data creation process, focusing on the mathematical formulation of the DC Optimal Power Flow (OPF) problem and the implementation of single-point attacks on the system.

3.1.1 DC Optimal Power Flow Formulation

The core of our simulation framework is built around a linearized DC OPF model implemented through two primary MATLAB functions: `lnzdOPF_fe` and `spaOPFsmlF`. The first function constructs the constraint matrices for the DC OPF problem, while the second orchestrates the overall simulation process including seasonal load modeling and attack implementation.

The `lnzdOPF_fe` function creates the equality constraints matrix, bound vectors, and right-hand side vector for the DC power flow equations. It takes three inputs:

- `Y`: The admittance matrix of the power network
- `P`: Vector of power values (negative values represent loads)
- `Pg`: Generator capacities (positive values)

The function organizes the optimization variables in the following structure:

- Bus active power: $\mathbf{P} = [P_0, P_1, \ldots, P_n]^T$
- Bus voltage angles: $\boldsymbol{\delta} = [\delta_0, \delta_1, \ldots, \delta_n]^T$
- Power flows: $\mathbf{F} = [P_{0,1}, \ldots, P_{n-1,n}, P_{1,0}, \ldots, P_{n,n-1}]^T$

The complete optimization vector combines these three components: $\mathbf{x} = [\mathbf{P}^T, \boldsymbol{\delta}^T, \mathbf{F}^T]^T$, where P_i represents the active power injection at bus i, δ_i is the voltage angle at bus i, and $P_{i,j}$ denotes the active power flow from bus i to bus j.

The lower and upper bounds for these variables are defined as:

$$\mathbf{lb} = [0_{1\times lenP}; -\pi \cdot 1_{1\times n}; -1000 \cdot 1_{1\times 2s\_flows}]^T \quad (1)$$

$$\mathbf{ub} = [Pg; \pi \cdot 1_{1\times n}; 1000 \cdot 1_{1\times 2s\_flows}]^T \quad (2)$$

The function then constructs the equality constraints matrix `Aeq` by iterating through all bus pairs. For each connected bus pair (i, j) (where $\text{imag}(Y_{j,i}) \neq 0$), it adds constraints that enforce:

1. The relationship between bus voltage angles and power flows:

$$P_{i,j} = \text{imag}(Y_{i,j}) \cdot (\delta_i - \delta_j) \tag{3}$$

2. Power balance at each bus:

$$P_i = \sum_{j \in \Omega_i} P_{i,j} \tag{4}$$

where Ω_i is the set of buses connected to bus i.

The right-hand side vector `beq` is initialized with the negative of the power values (to represent load balance equations), followed by zeros for the flow equations. Additional constraints are added for offline units (where $Pg = -100000$) to force their power output to zero. Finally, a constraint is added to set the angle of the first bus (slack bus) to zero.

3.1.2 Seasonal Load Modeling and Simulation Framework

The `spaOPFsmlF` function implements the overall simulation framework, which includes:

1. **Case Loading**: The function loads the specified IEEE 300-bus case using MATPOWER's functions and extracts the system parameters.

2. **Seasonal Load Modeling**: Four distinct seasonal load profiles (Winter, Spring, Summer, and Autumn) are created, each with 24-hour patterns that reflect typical daily load variations. These base profiles are then subjected to controlled randomization.

3. **DC OPF Setup**: For each of the 96 hours (24 hours × 4 seasons), the function: - Distributes the total system load to individual buses proportionally - Calls `lnzdOPF_fe` to create the constraints for each hour - Assembles these hourly constraints into larger matrices for each season and for the entire simulation period

4. **Generator Ramp Constraints**: The function adds constraints to limit the rate at which generators can change their output between consecutive hours:

$$-\text{RampRate}_i \cdot P_{g,i} \leq P_{g,i}^{t+1} - P_{g,i}^{t} \leq \text{RampRate}_i \cdot P_{g,i} \tag{5}$$

where RampRate_i is set to 0.5 for smaller generators and 0.3 for larger generators.

5. **Cost Matrix**: A linear cost function is defined based on the generator cost coefficients from the case data.

3.1.3 Attack Simulation

A key aspect of our methodology is the simulation of attacks on the power system. For each iteration, the function:

1. Solves the DC OPF under normal operating conditions using MATLAB's `linprog` function.

2. Randomly selects one generator and one of four attack types:

- **Type 1 (Ramp Rate Attack)**: Reduces the ramp rate capability of the targeted generator by a random factor between 0.01 and 1.

- **Type 2 (Upper Limit Attack)**: Reduces the maximum power output capability of the targeted generator.

- **Type 3 (Lower Limit Attack)**: Increases the minimum power output requirement of the targeted generator.

- **Type 4 (Cost Attack)**: Manipulates the cost coefficient of the targeted generator to a random value between half the minimum and 1.2 times the maximum of all generator costs.

3. Re-solves the DC OPF with the attack parameters applied to assess the impact on system operation.

4. Stores both normal and attack results for comparison and analysis.

3.1.4 Output Data Structure

The function generates a comprehensive output which is written to a CSV file. For each iteration, the output includes:

The normal operation data (`nrmF`) contains solution vectors with optimal power generation, voltage angles, and power flows for all hours and seasons, along with objective function values (`fvalTot`) representing minimum generation costs. It includes season indices (1-4) and zeros for attack indicators.

The attack scenario data (`attF`) contains equivalent solution vectors and objective values (`fvalTotATT`) for the compromised system, along with the same season indices plus the specific attack type (1-4) and targeted generator ID.

This dual structure enables direct comparison between normal and attacked states, capturing both the operational changes and economic impacts of different attack scenarios. This simulation process is repeated for the specified number of iterations, creating a comprehensive dataset that contains both normal operating conditions and various

attack scenarios across different seasonal patterns. The resulting data captures the economic impact of different attacks on the power system, which forms the basis for our machine learning model development.

3.2 Machine Learning Framework

After generating the synthetic power system data through the MATLAB simulation framework, we developed a comprehensive machine learning pipeline to detect anomalies in electricity market operations. Our approach focused on robust feature engineering, careful preprocessing, and systematic model evaluation to achieve high classification accuracy.

3.2.1 Data Processing and Feature Engineering

We began by transforming the raw simulation outputs into a structured dataset suitable for machine learning. The dataset contained 1920 records representing both normal operating conditions and various attack scenarios, with a total of 69 generators in the IEEE 300-bus system. Each record was labeled as either normal operation (0) or under attack (1), with additional metadata indicating the specific type of attack (1-4) and the targeted generator.

Domain-specific feature engineering was crucial to capture the complex patterns that distinguish normal operations from attacks. We created several categories of engineered features:

- **Session and temporal features**: These included day indices within each season to capture temporal patterns in electricity demand and supply.
- **OPF sensitivity indicators**: We developed features that quantify the sensitivity of objective function values to potential manipulations, such as normalized costs per unit load and season-specific variability metrics.
- **Rate-of-change indicators**: We engineered features tracking the rate of change of objective values, including differentials, acceleration metrics, and rolling statistics across multiple time windows. These metrics can identify anomalies in dynamic system behavior and operational transitions.
- **Operational boundary indicators**: We created features measuring proximity to operational bounds, including peak-to-trough ratios and quantile-based metrics within each season. These help detect when the system operates unusually close to or beyond typical constraints.
- **Economic efficiency metrics**: We developed efficiency scores comparing dispatch costs to seasonal averages and anomaly scores based on statistical deviations. These metrics help identify economically inefficient dispatch solutions that may indicate market manipulation.
- **Cross-session comparative features**: These metrics compared each session's behavior to the average across all sessions to identify unusual patterns.
- **Statistical anomaly indicators**: We created robust statistical metrics like median absolute deviations to detect outliers regardless of the specific anomaly mechanism.
- **Cumulative and trend features**: These tracked the evolution of system behavior over time to detect subtle manipulations that might otherwise go unnoticed in single-point analyses.
- **Lag-based features**: We incorporated autoregressive components by including lagged values and differences to capture temporal dependencies.

In total, we engineered over 30 features from the original simulation data to provide a rich set of indicators for the machine learning models to detect various attack patterns.

3.2.2 Feature Selection and Preprocessing

To improve model performance and interpretability, we implemented a two-stage feature selection process:

1. **Variance threshold filtering**: We first removed features with variance below 0.01, eliminating near-constant features that provide little discriminative power.

2. **Information-theoretic selection**: We applied mutual information scoring to identify the most relevant features for the classification task, selecting the top 100 features (or fewer if less than 100 remained after variance filtering).

The selected features were then scaled using a RobustScaler, which centers features around the median and scales according to the interquartile range. This approach ensured our models would be robust to outliers, which was essential given the nature of the anomaly detection task.

We applied a standard 80/20 train-test split to evaluate model performance, ensuring that the test set provided a realistic assessment of how well our models would generalize to unseen data.

3.2.3 Model Development and Evaluation

We developed and evaluated multiple machine learning models, each with distinct strengths for detecting different types of market anomalies:

1. **Extra Trees Classifier**: An ensemble of extremely randomized trees, offering high variance reduction through additional randomization compared to standard Random Forests.

2. **Gradient Boosting Classifier**: A sequential ensemble method that builds trees to correct the errors of previous trees, particularly effective for capturing complex patterns.

3. **Random Forest Classifier**: A bagging ensemble that combines multiple decision trees to reduce overfitting while maintaining predictive power.

4. **XGBoost Classifier**: An optimized gradient boosting implementation with enhanced regularization to prevent overfitting and improve generalization.

5. **Logistic Regression**: A linear model serving as a baseline and providing insights into the linear separability of normal and attack instances.

6. **Voting Classifier**: An ensemble meta-model that combines the predictions of the three best-performing base models to further improve accuracy and robustness.

For each model, we performed hyperparameter optimization using RandomizedSearchCV with stratified 5-fold cross-validation, optimizing for F1 score to balance precision and recall. This approach was critical for handling the class imbalance between normal operations and attacks.

The hyperparameter search spaces were extensive and model-specific. For tree-based models (Extra Trees, Gradient Boosting, Random Forest, XGBoost), we optimized parameters such as the number of estimators, maximum depth, minimum samples for splits and leaves, learning rates, and regularization terms. For Logistic Regression, we tuned the regularization strength, penalty type, and solver method.

3.2.4 Evaluation Metrics and Result Analysis

To comprehensively evaluate model performance, we employed multiple metrics:

- **Accuracy**: The proportion of correct predictions, providing an overall view of model performance.

- **F1 Score**: The harmonic mean of precision and recall, offering a balanced metric particularly suitable for imbalanced classes.

- **Confusion Matrix**: A tabulation of True Positives, False Positives, True Negatives, and False Negatives to understand the types of errors made by each model.

- **ROC-AUC**: The area under the Receiver Operating Characteristic curve, measuring the model's ability to discriminate between normal and attack conditions.

- **Feature Importance**: Analysis of which features contributed most to the classification decisions, providing insights into the most relevant indicators for detecting market anomalies.

For the full dataset analysis, we evaluated all models on the complete dataset containing all seasons. Additionally, we performed season-specific analyses by training and evaluating separate models for each season, allowing us to identify seasonal patterns in attack detectability.

Feature importance analysis was conducted for all tree-based models, enabling us to identify the most informative features for detecting different types of attacks. This analysis provided valuable insights into the mechanisms of how different attacks manifest in the power system data and which metrics are most useful for their detection.

4 Results and Discussion

Our experimental results demonstrate the effectiveness of machine learning techniques in detecting market anomalies in power systems across different seasonal conditions. We present performance metrics for both the full dataset model and season-specific models, highlighting the differences in detection capabilities under varying temporal contexts.

4.1 Overall Model Performance

We evaluated six different classification models on the complete dataset, which contained 1,920 records evenly split between normal operations and anomalous conditions. The best-performing model was a Voting Classifier ensemble combining Gradient Boosting, Random Forest, and XGBoost, achieving 92.45% accuracy and an F1 score of 0.9258.

Table 1 presents the performance metrics for all models trained on the full dataset, ranked by accuracy and F1 score.

Table 1: Model Performance on Full Dataset

Model	Accuracy	F1 Score
Voting Classifier	0.9245	0.9258
Gradient Boosting	0.9219	0.9227
Random Forest	0.9193	0.9215
XGBoost	0.9193	0.9211
Extra Trees	0.8646	0.8738
Logistic Regression	0.6823	0.7550

The confusion matrix for the best-performing model (Voting Classifier) is shown in Fig. 1. This model correctly identified 174 out of 190 normal operations (91.6% specificity) and 181 out of 194 anomalies (93.3% sensitivity), demonstrating a well-balanced classification performance.

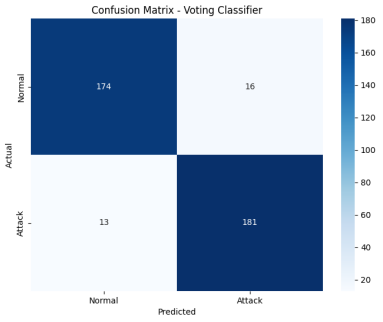


Figure 1: Confusion matrix for the Voting Classifier model on the full dataset.

The ROC curve for the Voting Classifier (Fig. 2) further confirms the model's strong discrimination capability, with an AUC (Area Under the Curve) of 0.9726. This high AUC value indicates that the model can effectively separate normal operations from anomalies across different classification thresholds.

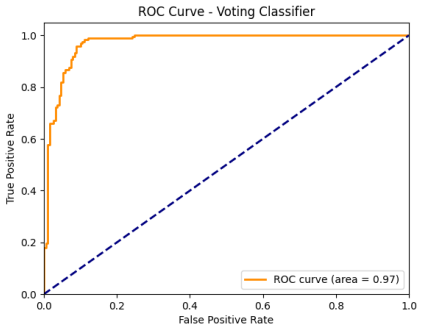


Figure 2: ROC curve for the Voting Classifier model on the full dataset, with AUC = 0.9726.

4.2 Seasonal Model Performance

To investigate the impact of seasonal variations on detection capability, we trained separate models for each of the four seasons represented in our dataset. Table 2 presents the best-performing models for each season along with their performance metrics.

The performance varied notably across different seasons, with Season 4 (Autumn) showing the highest accuracy and F1 score, while Season 1 (Winter) had the lowest performance metrics. This variation

Table 2: Best Model Performance by Season

Season	Best Model	Accuracy	F1 Score
1	GB*	0.8125	0.8302
2	GB*	0.8542	0.8627
3	GB*	0.8750	0.8800
4	GB*	0.8958	0.9057

*GB: Gradient Boosting

suggests that certain seasons may present more distinctive patterns between normal operations and anomalies, potentially due to differences in load profiles, generation patterns, or system constraints.

Fig. 3 shows the ROC curves for the best-performing models in each season. These curves demonstrate the trade-off between true positive rate and false positive rate across different classification thresholds.

4.3 Feature Importance Analysis

Analyzing the most influential features provides insights into which indicators are most valuable for detecting market anomalies. The feature importance analysis for the seasonal models is presented in Fig. 4. These plots highlight which features contributed most to the classification decisions in each seasonal context.

Across all models, the `fval_change` and `fval_diff_lag_1` features were consistently the most important across all models, highlighting the significance of temporal dynamics in detecting anomalies. Here, "fval" refers to the objective function value from the OPF solution, which represents the total generation cost of the system. The rate of change in this value is a powerful indicator of anomalies, as it captures unusual shifts in the economic dispatch of the system.

Interestingly, while the specific feature rankings varied between seasons, the general categories of important features remained consistent. This suggests that while the relative importance of specific indicators may shift with seasonal conditions, the fundamental patterns that distinguish normal from anomalous operation persist across different temporal contexts.

4.4 Comparative Analysis

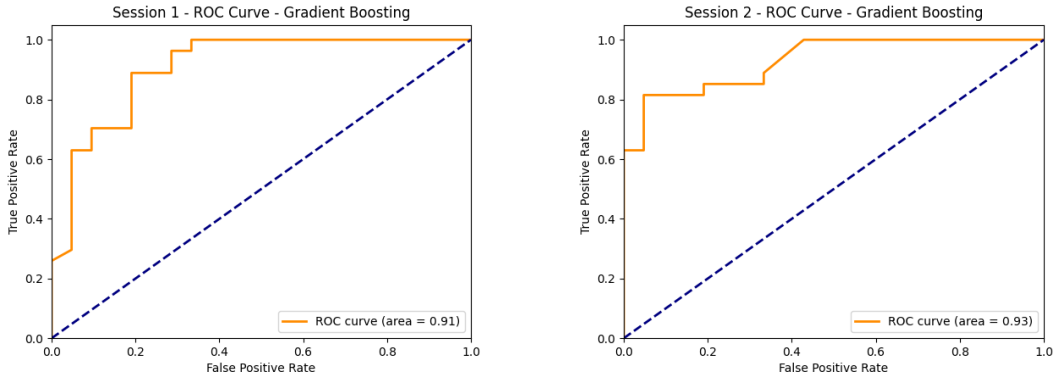
When comparing model performance across the full dataset versus seasonal subsets, we observed that the full dataset model achieved significantly higher accuracy (92.45% versus 81.25%–89.58% for seasonal models). This suggests that having a larger, more diverse training dataset improves the model's ability to generalize across different conditions.

An interesting finding is that while the Voting Classifier ensemble achieved the best performance on the full dataset (92.45% accuracy), Gradient Boosting consistently outperformed all other models in every seasonal subset. This paradox can be explained by several factors:

- **Reduced training data**: When training on seasonal subsets (480 samples vs. 1,920 for full dataset), Gradient Boosting's sequential learning may better adapt to limited data than ensemble methods that rely on model diversity.
- **Seasonal homogeneity**: Within individual seasons, data patterns may be more homogeneous, reducing the benefit of ensemble diversity. Gradient Boosting's iterative error correction may be more valuable when data exhibits consistent patterns.
- **Overfitting risks**: With smaller datasets, the Voting Classifier risks overfitting to its constituent models' training data. Gradient Boosting's regularization mechanisms may provide better generalization.

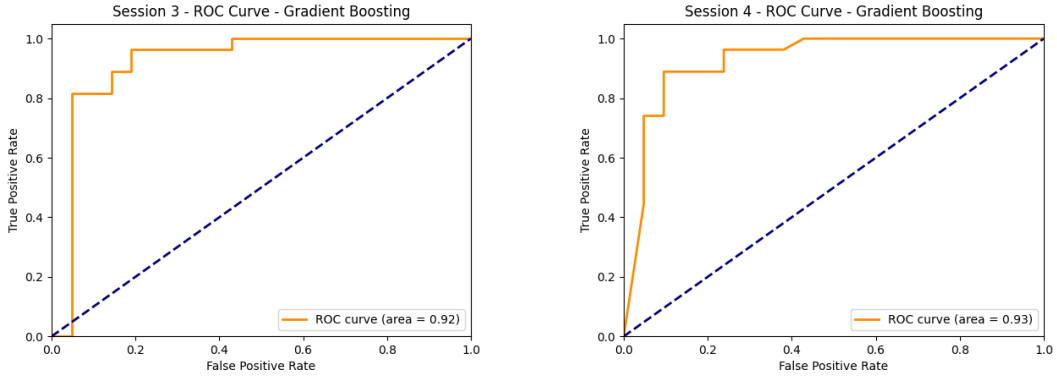
The seasonal analysis revealed important nuances:

- Detection accuracy varies substantially across seasons, with up to 8.33 percentage point differences between Season 4 (best) and Season 1 (worst).
- Gradient Boosting's dominance across all seasons suggests it captures fundamental patterns of power system anomalies more effectively than other models or combinations.
- F1 scores show more variation than accuracy, suggesting different precision-recall trade-offs by season.



(a) Season 1 - Gradient Boosting

(b) Season 2 - Gradient Boosting



(c) Season 3 - Gradient Boosting

(d) Season 4 - Gradient Boosting

Figure 3: ROC curves for the best seasonal models: (a) Season 1 - Gradient Boosting (AUC = 0.90274); (b) Season 2 - Gradient Boosting (AUC = 0.9295); (c) Season 3 - Gradient Boosting (AUC = 0.9206); (d) Season 4 - Gradient Boosting (AUC = 0.9286).

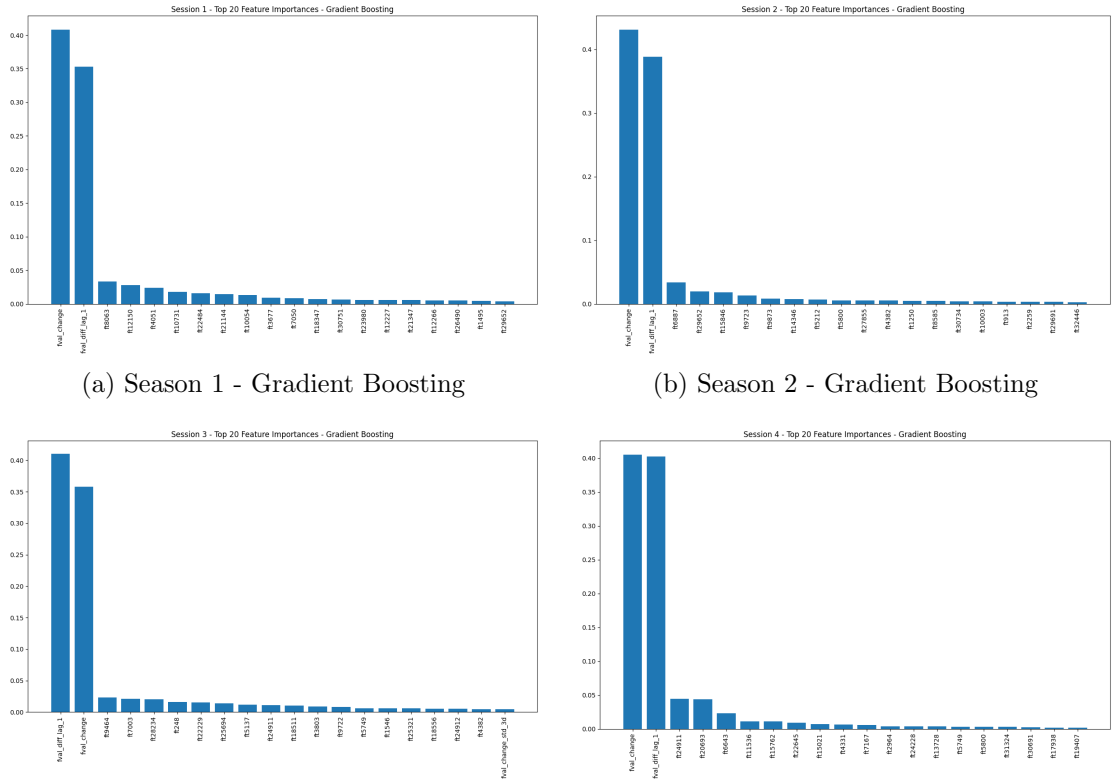
These findings highlight the importance of considering both model selection and temporal context when designing anomaly detection systems. While ensemble methods excel with large datasets, single models like Gradient Boosting may be more appropriate for season-specific or data-limited applications.

5 Conclusions

This work demonstrates that machine learning can effectively detect electricity market anomalies inspired by real-world incidents like the Ontario IESO phantom demand case. The strong performance of temporal features suggests that market manipulations fundamentally alter system dynamics in detectable ways, providing a foundation for practical monitoring systems.

The seasonal performance variations highlight an important trade-off: while ensemble methods excel with comprehensive datasets, single models like Gradient Boosting or any other individual model prove more robust for limited data scenarios. This suggests market operators should deploy different models based on available data and monitoring requirements.

Several promising research directions emerge from this work. Analyzing misclassified cases could reveal why certain attacks evade detection - possibly because they produce minimal changes in objective function values or target less influential



(a) Season 1 - Gradient Boosting

(b) Season 2 - Gradient Boosting

(c) Season 3 - Gradient Boosting

(d) Season 4 - Gradient Boosting

Figure 4: Feature importance plots for best seasonal models: (a) Season 1 - Gradient Boosting; (b) Season 2 - Gradient Boosting; (c) Season 3 - Gradient Boosting; (d) Season 4 - Gradient Boosting.

generators. Implementing unsupervised methods like isolation forests or autoencoders could identify novel attack patterns without requiring labeled data. Testing on different PGLib system sizes would evaluate scalability, while generator-specific models could capture unique vulnerability patterns.

Real-time implementation presents additional challenges, requiring online learning capabilities to adapt to evolving threats while maintaining high accuracy. As electricity markets grow more complex with renewable integration and demand response programs, robust anomaly detection systems become increasingly critical for protecting market integrity and consumer interests.

References

[1] M. Sharp, "Ontario's electricity operator kept quiet about phantom demand that cost customers millions," *National Observer*, Dec. 23, 2019. [Online]. Available: `https://www.nationalobserver.com/2019/12/23/news/ontarios-electricity-operator-kept-quiet-about-phantom-demand-that-cost-customers`

[2] S. Babaeinejadsarookolaee, A. Birchfield, R. D. Christie, C. Coffrin, C. DeMarco, R. Diao, M. Ferris, S. Fliscounakis, S. Greene, R. Huang, C. Josz, R. Korab, B. Lesieutre, J. Maeght, D. K. Molzahn, T. J. Overbye, P. Panciatici, B. Park, J. Snodgrass, and R. Zimmerman, "The Power Grid Library for Benchmarking AC Optimal Power Flow Al-

gorithms," 2019. [Online]. Available: https://github.com/power-grid-lib/pglib-opf

[3] M. Kardi, T. AlSkaif, B. Tekinerdogan and J. P. S. Catalão, "Anomaly Detection in Electricity Consumption Data using Deep Learning," 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Bari, Italy, 2021, pp. 1-6, doi: 10.1109/EEEIC/ICPSEurope51590.2021.9584650. keywords: Deep learning;Wind speed;Neural networks;Europe;Humidity;Predictive models;Feature extraction;LSTM autoencoder;Deep Learning;Anomaly detection;Electricity consumption;Anomalous consumption.

[4] H. Guo, Q. Chen, Y. Gu, M. Shahidehpour, Q. Xia and C. Kang, "A Data-Driven Pattern Extraction Method for Analyzing Bidding Behaviors in Power Markets," in IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3509-3521, July 2020, doi: 10.1109/TSG.2019.2962842. keywords: Power markets;Clustering methods;Generators;Standardization;Power systems;Data mining;Data-driven power market analyses;bidding behavior;pattern extraction

[5] M. Sun, L. Ren and N. -y. Chiang, "Data-Driven Probabilistic Anomaly Detection for Electricity Market under Cyber Attacks," 2021 American Control Conference (ACC), New Orleans, LA, USA, 2021, pp. 4586-4591, doi: 10.23919/ACC50511.2021.9482640. keywords: Electric potential;Sensitivity analysis;Electricity supply industry;Probabilistic logic;Smart grids;Numerical models;Reliability