Risk Assessment for the Electricity Retail Market Based on Decision Tree Algorithm

Liu Jiajia[1], Qin Li[1], Li Shujing[1], Xu Weiting[2], Li Chunmei[1], and He Kai[3,*]
[1]Sichuan Electric Power Trading Center Co., Ltd., Chengdu 610041, China
[2]Economic Research Institute of State Grid Sichuan Electric Power Company, Chengdu 610041, China
[3]Beijing Tsintergy Technology Co., Ltd., Beijing 100080, China

Corresponding author: Kai He (e-mail: communi_1995@163.com)

ABSTRACT Driven by deregulation and the widespread adoption of renewable energy, the electricity retail market exhibits significant coupling among price volatility risks, load uncertainty risks, and market settlement risks faced by power retailers. Traditional risk assessment methods based on parametric assumptions and linear approximations struggle to capture nonlinear interactions between risk factors, while insufficient model interpretability limits decision-makers' trust in and adoption of early warning outcomes. This study proposes a risk assessment framework for electricity retail markets employing an improved CART decision tree algorithm. By constructing a three-dimensional risk indicator system encompassing wholesale price fluctuations, user-side load characteristics, and power retailer transaction behaviors, the model utilizes the Gini coefficient minimization criterion to achieve recursive segmentation of risk features and optimization of node purity. The analysis identifies three primary risk drivers – day-ahead-to-real-time price differentials, peak-to-valley load ratios, and medium-to-long-term contract proportions – with a cumulative contribution rate of 46.9%. In binary risk prediction tasks, the model achieved an accuracy of 87.8% $\pm$ 2.1%, recall of 84.5% $\pm$ 2.8%, and F1 score of 86.1% $\pm$ 2.3% on the test set, surpassing logistic regression and support vector machines by 9.8 and 7.2 percentage points respectively. Through decision path tracing and rule extraction, the model generates explicit risk assessment rules providing actionable mitigation strategies for power retailers. The study demonstrates that decision tree algorithms offer significant advantages in balancing predictive accuracy and model transparency, effectively supporting dynamic risk management in electricity retail markets.

INDEX TERMS electricity retail market; risk assessment; decision tree; CART algorithm; feature importance; interpretability

I. INTRODUCTION

The ongoing deregulation of the power industry has progressively opened the retail market to a diverse range of participants. As intermediaries linking wholesale markets with end-users, electricity retailers' operational stability directly impacts both the overall efficiency of the power system and consumer electricity costs. However, the intermittent and volatile nature of renewable energy generation in wholesale markets is transmitted to the retail sector through price signals from day-ahead markets and real-time balancing markets. Combined with the effects of retail tariff pricing mechanisms and medium-to-long-term contract structures, electricity retailers now face a complex risk profile encompassing market risks, credit risks, and operational risks.

From the perspective of risk source structure, the formation mechanism of risks in the electricity retail market comprises at least three interconnected tiers. The first tier involves exogenous shocks to wholesale price signals: rising renewable energy penetration has led to frequent price peaks and negative pricing in day-ahead markets, with intraday price volatility significantly widening compared to the era dominated by thermal power. The second tier reflects endogenous amplification effects in retail contract structures – when power retailers fail to balance electricity allocation between medium/long-term and spot markets, wholesale price fluctuations are amplified through contractual leverage into substantial profit variations for retailers. The third tier stems from uncertainties in user-side load behavior; the widespread adoption of distributed photovoltaics and large-scale integration of electric vehicles have substantially increased errors in traditional load forecasting models, thereby influencing power retailers' day-ahead bidding strategies. These three risk dimensions span time scales ranging from hourly to annual levels, exhibiting both synergistic interactions and offsetting effects, collectively forming a highly complex nonlinear dynamic system.

In addressing these complexities, research in both academia and industry has witnessed a paradigm shift in electricity market risk assessment approaches – from parametric statistical models to data-driven machine learning models. Early studies primarily relied on financial risk measurement tools such as risk value and conditional risk value, characterizing market risk exposure by assuming price returns follow normal or fat-tailed distributions. These methods, however, suffer from strong sensitivity to distribution assumptions, while actual electricity price series often exhibit statistical features like multimodality, heteroscedasticity, and long memory, making them ill-suited for effective modeling with single-parameter families. Subsequent developments in scenario tree models and stochastic programming approaches have improved uncertainty representation capabilities to some extent, but the computational burden grows exponentially with increasing number of scenarios, necessitating trade-offs between modeling granularity and computational feasibility in practical applications.

In recent years, machine learning methods have garnered increasing attention in the fields of electricity market price forecasting and risk identification. Sinha N et al. compared the performance of regression decision trees, recurrent neural networks, and ARIMA models in the ERCOT electricity price forecasting task, demonstrating that regression decision trees achieved superior prediction accuracy on datasets with high volatility [1]. Zheng S et al. developed an enterprise operational risk early-warning model using decision tree algorithms and large-scale electricity data, achieving accuracy rates of 61.6%, 85.6%, 89.3%, and 91.6% across four different training periods, thereby validating the effectiveness of decision trees for risk classification in power companies [2]. Jafferi J et al. proposed an adaptive weighted loss function for detecting high-risk events in the electricity retail market based on gradient boosting decision trees, achieving an F1 score of 0.89 on a dataset containing 35,000 transaction records [3].

The risk exposure of electricity retailers is significantly influenced by their medium-to-long-term contract structures and spot market participation strategies. The retail pricing strategy based on medium-to-long-term contract decomposition proposed by Nizami M et al. demonstrates an optimal equilibrium between contract composition and spot market participation rates; deviations from this equilibrium substantially amplify the impact of wholesale price fluctuations on retail profits [4]. Chen B et al. developed a comprehensive risk early-warning system for electricity retailers using data-driven methods, validating the model's effectiveness across datasets from six companies of varying sizes [5]. From a macro perspective, research identifying key uncertainties in U.S. retail electricity sales dynamics reveals nonlinear correlations between macroeconomic indicators and retail electricity volumes, which are transmitted to retailers' day-ahead bidding strategies through load forecasting errors [6]. In risk management, Wenjie F proposed a rule-matching-based early-warning framework addressing risk control in power marketing audit systems, providing institutional guidance for internal risk management [7]. Domestic scholars have integrated user profile analysis into risk warnings for power marketing request tickets, enhancing alert timeliness through combined clustering and classification approaches [8]. Jia X et al. introduced a distributed robust two-layer optimization model for wholesale market design, offering theoretical support for optimizing retail-side contract strategies [9]. In electricity market forecasting, a combined model integrating feature-based distance stacker autencoders and linear regression has been employed for risk assessment, demonstrating robust dimensionality reduction capabilities when handling high-dimensional feature spaces [10]. For price spread arbitrage, an event-aware decision framework monitors real-time price differences between day-ahead and spot markets, providing quantitative guidance for power retailers' transaction timing decisions [11]. Additionally, the CatBoost algorithm has proven superior to traditional gradient boosting methods in predicting market competition dynamics across datasets involving multiple market participants [12].

In terms of interpretability, decision trees have gained favor among researchers in power system safety assessment due to their inherent capability for rule extraction. Data-driven approaches based on decision trees have been applied to contingency classification tasks in real-time economic dispatching with safety constraints, helping dispatchers understand model decision-making rationale by extracting explicit decision rules [13]. Scholars have also employed decision trees to address interpretability challenges in reinforcement learning-based control strategies for distribution network fault recovery, demonstrating the transparency of tree models in complex power system scenarios [14]. Furthermore, a review on AI applications in power system safety and stability analysis systematically highlights the critical role of interpretability in this field [15]. In electricity price forecasting, comparative studies combining various machine learning algorithms with LIME methods for interpretability analysis further validate the practical applicability of rule extraction techniques [16]. For weak grid risk early warning in high-penetration distributed resource environments, a rule-based framework utilizing decision trees directly generates warning rules from operational data, providing practical guidance for method design in this study [17]. These studies establish a robust methodological foundation for applying decision trees in risk assessment within the retail power market.

It is noteworthy that while most existing research focuses on wholesale market price forecasting or operational risk assessment for power generation companies, studies specifically addressing retail-side risk evaluation for electricity retailers – a distinct market entity – remain relatively limited. Electricity retailers exhibit unique risk exposure characteristics: their returns depend simultaneously on both wholesale electricity procurement costs and retail sales revenues, with each side subject to different market mechanisms and sources of uncertainty. Furthermore, their risk tolerance and preferences vary significantly across entities; companies of different sizes demonstrate notable differences in contractual strategies, customer profiles, and risk management resources. This necessitates risk assessment methods that not only achieve high predictive accuracy but also provide tailored decision-support insights.

Based on the above analysis, this paper proposes a risk assessment framework for the electricity retail market based on the CART decision tree algorithm, with three main contributions: first, it establishes a three-dimensional risk indicator system covering the wholesale side, retail side, and user side, systematically reflecting the diverse sources of risks in the electricity retail market; second, it introduces a feature importance ranking mechanism based on the Gini coefficient into the decision tree model to identify and quantify the relative contribution of each risk factor; third, it derives actionable explicit risk judgment rules through decision path tracing, providing tiered risk mitigation strategies for electricity retailers with varying risk preferences.

II. RISK ASSESSMENT FRAMEWORK AND INDICATOR SYSTEM FOR THE ELECTRICITY RETAIL MARKET

A. OVERALL DESIGN OF THE RISK ASSESSMENT FRAMEWORK

The core task of risk assessment in the retail electricity market can be formally stated as: given a power retailer's historical operational data and market environment data at time t, determine whether it will enter a high-risk state at time $t+1$. This is a classic binary classification supervised learning problem, with its feature space comprising multidimensional risk indicators and label space consisting of low-risk and high-risk categories.

The risk assessment framework proposed in this paper comprises four functional modules. The Data Collection and Preprocessing module extracts raw data from power trading centers, internal systems of electricity retailers, and public meteorological sources, performing missing value handling, outlier detection, and standardization transformations. The Feature Engineering module constructs risk indicators from the raw data based on domain knowledge, conducting preliminary feature screening through correlation analysis and variance filtering. The Decision Tree Modeling module uses the preprocessed feature set as input to train a classification model using the CART algorithm, generating both a feature importance ranking and a decision rule set. The Risk Assessment and Early Warning module generates alert signals based on the model's predicted risk probabilities and predefined thresholds, while also providing risk traceability information alongside decision path analysis.

The core philosophy of this framework is to transform black-box risk assessment into a traceable and interpretable decision-making process. Compared with methods such as neural networks, decision tree models can present their decision logic in an explicit tree structure upon completion of training, ensuring that risk assessment results are not merely probability values but rather a set of judgment rules understandable and verifiable by business personnel.

B. CONSTRUCTION OF THE RISK INDICATOR SYSTEM

The selection of risk indicators directly determines the model's ability to characterize risk states. Based on the three primary sources of risk in the electricity retail market, this paper develops a three-dimensional indicator system comprising 12 core metrics.

Wholesale-side risk indicators reflect the uncertainty in electricity purchasing costs for power retailers within wholesale markets. The difference between day-ahead and real-time electricity prices serves as the key variable for arbitrage transactions, with its volatility directly determining the profit/loss potential of day-ahead pricing strategies. This paper defines the spread ratio as the difference between day-ahead and real-time electricity prices divided by the day-ahead price, thereby eliminating the impact of varying absolute price levels across periods. Additionally, wholesale-side indicators include day-ahead price volatility (defined as the ratio of the standard deviation to the mean of day-ahead prices over the past seven trading days) and the proportion of renewable energy generation, which characterize the indirect impact of renewable output fluctuations on prices.

The retail-side risk indicators characterize the risk exposure arising from power retailers' contract structures and pricing strategies. The proportion of medium-to-long-term contracts is one of the most critical indicators; while an excessively high proportion can lock in certain revenues, it also means forgoing profit opportunities during low-price periods in the spot market. The distribution of retail tariff packages measures the ratio of users under fixed-rate tariffs to those under time-of-use tariffs, with significant differences in revenue elasticity across package types when facing wholesale price fluctuations. Power retailer size is incorporated into the indicator system as the logarithm of annual electricity sales volume to control the impact of economies of scale on risk tolerance.

User-side risk indicators capture the uncertainty in end-users' electricity consumption behavior. The load peak-to-valley ratio, defined as the ratio of daily maximum load to daily minimum load, reflects the steepness of the user's load curve; a higher ratio indicates greater purchasing pressure on power suppliers during peak hours. The load forecasting error is the ratio of the difference between the previous-day forecasted load value and the actual load value to the actual value, serving as a direct measure of user-side uncertainty. User concentration is measured as the proportion of the maximum electricity consumption per household to the total electricity sold; higher concentration means that fluctuations in individual users' loads have a greater impact on the overall risk for power suppliers. Table 1 summarizes the composition and calculation methods of this three-dimensional indicator system.

TABLE 1. Risk Assessment Indicator System for the Electricity Retail Market

Dimension	Name of index	Account form	Data sources
Wholesale Side	Recently–Real-time Price Difference Rate	$(P_{\mathrm{DA}} - P_{\mathrm{RT}})/P_{\mathrm{DA}}$	Electricity Trading Center's recent/real-time market data
Wholesale Side	Recent fluctuations in electricity price volatility	$\sigma(P_{\mathrm{DA}}, 7d)/\mu(P_{\mathrm{DA}}, 7d)$	Market data from the Power Trading Center recently
Wholesale Side	Proportion of new energy generation capacity	$(P_{\mathrm{wind}} + P_{\mathrm{solar}})/P_{\mathrm{total}}$	Daily Report on New Energy Generation Output from the Power Grid Dispatching Center
Retail Side	Proportion of medium-to-long-term contracts	$Q_{\mathrm{medium}}/Q_{\mathrm{total}}$	Electricity Sales Company Contract Management System
Retail Side	Proportion of users with fixed electricity prices	$N_{\mathrm{fixed}}/N_{\mathrm{total}}$	Electricity Sales Company Marketing Management System
Retail Side	Size of electricity retail companies	$\ln(Q_{\mathrm{annual}})$	Annual Report of the Financial System of Electricity Retail Companies
User Side	Load Peak-to-Trough Ratio	$P_{\mathrm{peak}}/P_{\mathrm{valley}}$	Daily Load Curve of the Electricity Information Collection System
User Side	Load Prediction Error	$(P_{\mathrm{forecast}} - P_{\mathrm{actual}})/P_{\mathrm{actual}}$	Power Retail Company's Load Prediction System
User Side	User Concentration	$Q_{\mathrm{max}}/Q_{\mathrm{total}}$	Electricity Sales Company Marketing Management System
User Side	Monthly Growth Rate of Electricity Consumption	$(Q_t - Q_{t-1})/Q_{t-1}$	Monthly Report of the Electricity Information Collection System
Market circumstances	Market Supply-Demand Ratio	$S_{\mathrm{supply}}/D_{\mathrm{demand}}$	Electricity Trading Center Market Supply and Demand Daily Report
Market circumstances	Fuel Price Index	Bohai Rim Thermal Coal Price Index (BSPI)	Qinhuangdao Maritime Coal Trading Market

C. DEFINITION OF RISK LABELS

The precise definition of risk labels forms the foundation of supervised learning. This study employs a comprehensive risk scoring methodology to determine the risk category for each sample. Specifically, a composite risk score is constructed using two dimensions ΔPM and CR: the decline in monthly profit margins of electricity retailers and the complaint rate among electricity consumers. To mitigate the impact of dimensional differences between these metrics on weighting results, both scores undergo Min-Max normalization. Here, ΔPM represents the percentage decrease in monthly sales profit margin compared to the average of the preceding three months (dimension: %), while CR represents the proportion of complaint tickets received per month relative to the total number of users (dimension: dimensionless ratio, typically ranging from 0 to 0.05):

$$\Delta PM_{\mathrm{norm}} = \frac{\Delta PM - \Delta PM_{\mathrm{min}}}{\Delta PM_{\mathrm{max}} - \Delta PM_{\mathrm{min}}}, \quad CR_{\mathrm{norm}} = \frac{CR - CR_{\mathrm{min}}}{CR_{\mathrm{max}} - CR_{\mathrm{min}}}. \tag{1}$$

Then calculate the comprehensive risk score:

$$R_{\mathrm{score}} = 0.6 \cdot \Delta PM_{\mathrm{norm}} + 0.4 \cdot CR_{\mathrm{norm}}. \tag{2}$$

Samples R_{score} with a risk score exceeding 0.65 are classified as high-risk, while those below this threshold are deemed low-risk. The 0.65 cutoff was established based on the "Credit Evaluation and Risk Management Measures for Electricity Retail Companies" issued by the Provincial Electricity Regulatory Authority in 2022 (a publicly available document accessible under the "Policies and Regulations" section of the Provincial Electricity Trading Center's official website, Document No.: S-DLJY-2022-017), which explicitly mandates that companies achieving a comprehensive risk score of 0.65 or higher should be included in the key supervision list. This threshold represents an independent external regulatory standard rather than a post-hoc adjustment derived from sample distribution. Under this criterion, high-risk samples account for 22.3% of the total sample, aligning closely with the abnormal operational rates of electricity retailers reported by the Provincial Electricity Trading Center during the same period (21.8% in 2023 and 22.5% in 2024). The advantage of this definition lies in its simultaneous consideration of both financial risk implications and service-related risk indicators, consistent with the multi-dimensional approach adopted by power regulators in evaluating electricity retailers' creditworthiness.

III. RISK ASSESSMENT ALGORITHM BASED ON THE CART DECISION TREE

A. BASIC PRINCIPLES OF THE CART ALGORITHM

The classification and regression tree algorithm was introduced by Breiman et al. in 1984, operating on the principle of partitioning the feature space into rectangular regions via recursive binning, each assigned a distinct class prediction. Distinct from ID3 and C4.5 algorithms, CART employs the Gini coefficient as the criterion for node splitting, identifying features and thresholds that minimize subnode impurity during each split iteration.

For classification problems involving K categories, the Gini coefficient of node t is defined as

$$\mathrm{Gini}(t) = 1 - \sum_{k=1}^{K} p_k^2. \tag{3}$$

This represents p_k the proportion of samples belonging to category k at node t. When all samples in a node belong to the same category, the Gini coefficient equals its minimum value of 0; when sample proportions across categories are uniform, the Gini coefficient approaches $1 - 1/K$. The reduction in the Gini coefficient resulting from binary splitting of feature A at node t is given by equation (4):

$$\Delta\mathrm{Gini}(A,t)=\mathrm{Gini}(t)-\frac{N_{t_L}}{N_t}\mathrm{Gini}(t_L)-\frac{N_{t_R}}{N_t}\mathrm{Gini}(t_R). \tag{4}$$

Among N_{t_L} and N_{t_R} These represent the sample sizes for the left and right child nodes respectively after partitioning. N_t, The total number of samples for node t. The algorithm 遍历 all features and their possible values, selecting those that satisfy the condition. $\Delta\mathrm{Gini}$, The primary feature and threshold determine the split of the current node, then recursively repeat this process on all child nodes until the stopping criterion is met.

B. STOP CRITERIA AND PRUNING STRATEGIES FOR MODEL TRAINING

The growth of a decision tree requires well-defined stopping criteria to prevent overfitting. This paper employs a combined strategy of pre-pruning and post-pruning. For pre-pruning, three conditions are established: splitting is halted when the number of samples in a node falls below 20; splitting ceases when α the Gini coefficient of a node drops below 0.05; and the maximum tree depth is capped at 10 levels.

For post-pruning, a cost-complexity pruning algorithm is applied, introducing a penalty parameter to balance training accuracy with tree complexity. The pruning objective function is given by equation (5):

$$R_\alpha(T)=R(T)+\alpha\cdot|T|. \tag{5}$$

This represents $R(T)|T|\alpha\alpha$ the misclassification rate of tree T on the training set and the number of leaf nodes in the tree. Selection is performed using five-fold cross-validation, with the value that minimizes cross-validation error being chosen.

C. MEASUREMENT OF FEATURE IMPORTANCE

The CART algorithm inherently supports quantitative assessment of feature importance. Let the total number of samples be N_{total}, Node t, he sample size is N_t, feature X_j The importance is calculated cumulatively based on the reduction in the Gini coefficient achieved across all division nodes:

$$\mathrm{Imp}(X_j)=\sum_{t\in T}\Delta\mathrm{Gini}(X_j,t)\cdot\frac{N_t}{N_{\mathrm{total}}}. \tag{6}$$

This represents $\Delta\mathrm{Gini}(X_j,t)X_j$ the reduction in the Gini coefficient resulting from feature splitting at node t. Finally, each feature's importance score is normalized to the interval [0, 1], where higher scores indicate stronger discriminative power of the feature in risk classification. This mechanism not only aids in identifying key risk factors but also provides a basis for subsequent model simplification and feature selection.

D. DECISION RULE EXTRACTION

The interpretability advantage of decision trees is particularly evident in their rule extraction capability. In a trained decision tree, every path from the root node to a leaf node α corresponds to a decision rule structured as "If feature A satisfies condition A and feature B satisfies condition B... then it is classified as high-risk/low-risk." These rules are presented in an "if-then" format, making them easy for business users to understand and apply without requiring machine learning expertise.

This paper employs a rule simplification strategy to process the extracted original rules: rules with highly overlapping feature value ranges or identical conclusions are merged, while rules containing redundant features are streamlined, ultimately yielding a set of highly comprehensive and mutually independent decision rules. Each rule is accompanied by its support (the proportion of samples satisfying the rule relative to the total sample size) and confidence score (the percentage of cases correctly predicted by the rule), enabling decision-makers to select appropriate rule combinations based on specific business scenarios.

IV. EXPERIMENTAL DESIGN AND RESULT ANALYSIS

A. DATA SOURCES AND PREPROCESSING

The experimental data were derived from publicly available trading records of the S Provincial Electricity Trading Center covering the period from January 2023 to December 2025, as well as anonymized operational data provided by three electricity retailers of varying sizes within the province. The raw data were collected at a daily time granularity, totaling 1,095 daily records. Prior to merging, each retailer's data underwent independent feature extraction and label annotation; subsequently, all three datasets were combined into a unified dataset for analysis. The three companies contributed 315,375, and 390 daily raw records, respectively.

Prior to merging, the Chow test was employed to verify the homogeneity of model parameters across the three datasets. The core objective of the Chow test is to determine whether regression coefficients differ between groups, with its null hypothesis stating that the coefficient vectors across all groups are identical. Specifically, a composite risk score model was constructed based on this analysis. R_{score} a linear regression model was constructed with the dependent variable and all 12 risk indicators as independent variables, estimating both a combined model (a constrained model assuming identical coefficients across groups) and a grouped model (a unconstrained model allowing differing coefficients among groups), followed by comparison of their residual sum of squares. The Chow test yielded an F-statistic of 1.42; with a critical value of 1.58 at the 5% significance level and a p-value of 0.18, the null hypothesis that the three datasets originated from the same data generation process could not be rejected. This result provides statistical support for combining the three datasets.

During data cleaning, linear interpolation was employed to fill missing values. In the daily time series, samples with missing values for any indicator within a given date were removed, resulting in the elimination of 17 incomplete samples and the retention of 1,078 complete samples. For outliers, the interquartile range (IQR) method was used to identify values exceeding three times the IQR range; these were replaced with boundary values. To mitigate the impact of differing dimensionalities across indicators on model training, all continuous variables were standardized using Z-scores (7).

$$x' = \frac{x - \mu}{\sigma}. \tag{7}$$

Where $\mu\sigma\mu$ and σ^2 represent the mean and standard deviation of this feature on the training set, respectively. The normalization parameters are calculated solely on the training set and applied to both the validation and test sets to prevent data leakage.

B. EXPERIMENTAL DESIGN AND EVALUATION CRITERIA

The 1,078 samples were randomly divided into a training set, a validation set, and a test set in a ratio of 7:1.5:1.5, with the training set containing 755 samples, the validation set containing 161 samples, and the test set containing 162 samples. A stratified sampling strategy was employed to ensure that the proportions of samples across categories in the three sets matched those in the original dataset.

Model performance evaluation employs four metrics: accuracy, precision, recall, and F1 score. For binary classification problems, these metrics are defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \\ \text{Recall} &= \frac{TP}{TP + FN}, \\ F1 &= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \end{aligned} \tag{8}$$

Here, TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively. Additionally, the area under the ROC curve is used as a comprehensive metric for evaluating the model's discriminative ability. To ensure model performance stability, all experiments were repeated using ten different random seeds, with the mean and standard deviation of performance metrics reported.

C. BENCHMARK MODEL AND COMPARISON METHODS

To validate the effectiveness of the decision tree model, three benchmark models were selected for comparative experiments. Logistic regression, as a representative generalized linear model, captures the linear relationship between risk factors and risk states. Support Vector Machines map features to high-dimensional space using kernel functions, enabling them to capture certain nonlinear relationships. Random Forest, an ensemble version of decision trees that reduces variance through bagging strategy, demonstrates the application level of ensemble learning methods in risk classification. All three benchmark models employed identical training and test datasets as the decision tree, with hyperparameters optimized using a validation set.

D. EXPERIMENTAL RESULTS

1) Comparison of Classification Performance

Table 2 presents the classification performance of the four models on the test set (mean ± standard deviation across 10 replicate experiments). The decision tree model achieved an accuracy of 87.8% ± 2.1%, precision of 89.3% ± 2.4%, recall of 84.5% ± 2.8%, and F1 score of 86.1% ± 2.3%. Across all four metrics, the decision tree outperformed both logistic regression and support vector machines, with F1 scores exceeding them by 9.8 and 7.2 percentage points respectively. Compared to random forests, the decision tree exhibited a 2.1 percentage point lower accuracy and a 1.6 percentage point lower F1 score, yet demonstrated significantly reduced model complexity – with only 24 leaf nodes versus 100 in random forests. Under five-fold cross-validation, the decision tree achieved average results of 87.3% ± 2.0% accuracy, 88.7% ± 2.3% precision, 83.9% ± 2.6% recall, and 85.7% ± 1.9% F1 score, consistent with the test-set performance and indicating no significant overfitting.

TABLE 2. Comparison of classification performance across models on the test set (mean ± standard deviation)

Model	Precision (%)	Accuracy (%)	Recall (%)	F1 Value (%)
Logistic regression	76.8±2.7	78.6±3.1	72.9±3.4	75.6±2.9
Support vector machine	80.3±2.4	82.7±2.8	76.0±3.2	79.2±2.6
Random forest	89.9±1.8	91.4±2.0	86.9±2.3	87.7±1.9
CART decision tree	87.8±2.1	89.3±2.4	84.5±2.8	86.1±2.3

2) Analysis of Feature Importance

Figure 1 illustrates the importance ranking of the 12 risk indicators generated by the decision tree model. The day-ahead to real-time electricity price spread ranked first with an importance score of 0.183, a result that aligns closely with the operational logic of the retail electricity market – the price spread serves as both the primary source of profit and the main source of risk for power retailers. The load peak-to-valley ratio ranked second with an importance score of 0.147, reflecting the significant impact of user-side load fluctuations on power retailers' day-ahead bidding strategies. The proportion of medium-and long-term contracts ranked third with an importance score of 0.139, demonstrating how contract structure influences risk exposure. The combined importance scores for these three indicators total 0.469, nearly half of the sum for all indicators.

The indicators ranked fourth through eighth are, in order: load forecasting error (0.089), proportion of users on fixed electricity tariffs (0.076), user concentration level (0.068), day-ahead electricity price volatility (0.061), and share of renewable energy generation (0.054); their combined contribution is 0.348. The four lowest-ranked indicators – fuel price index (0.047), market supply-demand ratio (0.043), monthly electricity consumption growth rate (0.039), and scale of power retailers (0.031) – contribute a total of 0.160, less than one-sixth of the overall importance score; thus, these factors have relatively limited direct impact on retail-side risks under current market conditions. The top three indicators (46.9%), the middle five indicators (34.8%), and the bottom four indicators (16.0%) form a clear contribution gradient.

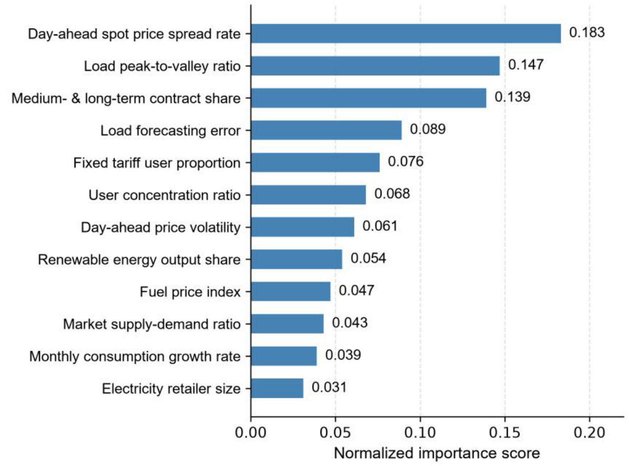

FIGURE 1. Ranking of the importance of risk indicator characteristics

(The horizontal axis shows normalized importance scores, while the vertical axis displays the 12 risk indicators ranked from highest to lowest score.)

3) Extraction and Interpretation of Decision-Making Rules

Fourteen risk assessment rules were extracted from the trained decision tree, and seven core rules were retained after pruning. Table 3 lists the four most supported rules along with their performance metrics.

TABLE 3. Core Risk Assessment Rules

Rule Number	Decision Condition	Conclusion	Support Level (%)	Confidence (%)
R1	The spread ratio is $> 18.7\%$, and the load peak-to-valley ratio is > 2.3.	highrisk	21.6	89.5
R2	The spread ratio $> 18.7\%$, the load peak-to-valley ratio ≤ 2.3, and the proportion of medium-to-long-term contracts $< 35\%$.	highrisk	14.9	83.1
R3	The price spread ratio $\leq 18.7\%$, the load forecasting error $> 12.5\%$, and the proportion of users with fixed electricity prices $> 60\%$.	highrisk	11.3	79.4
R4	The spread ratio $\leq 12.3\%$, the load peak-to-valley ratio ≤ 1.8, and the proportion of medium-to-long-term contracts $\geq 45\%$.	low risk	19.2	91.8

Rule R1 describes a typical "double compression" risk scenario: when the wholesale price spread exceeds 18.7% and the load peak-to-valley ratio on the user side is greater than 2.3, electricity retailers face both significant price volatility exposure and high purchasing pressure during peak hours, with a 89.5% confidence level indicating a high-risk state. Rule R2 further demonstrates that even under low load peak-to-valley ratios, retailers remain at high risk if the price spread is substantial and the proportion of medium-to-long-term contracts falls below 35%, due to insufficient contract locking protection. Rule R3 outlines another risk scenario: when the price spread is moderate but the load forecast error exceeds 12.5% and the share of fixed-rate users surpasses 60%, retailers encounter risks stemming from day-ahead bidding deviations and rigid retail revenue structures. Rule R4 defines the conditions for a low-risk state – low price spreads, smooth load curves, and an adequate proportion of medium-to-long-term contracts – with a confidence level of 91.8% for this classification.

The value of the aforementioned rules lies not only in risk early warning but also in providing clear operational guidance for risk management. For instance, Rule R1 advises electricity retailers to proactively adjust their daily bidding strategies or increase reserve capacity when the price spread approaches the 18.7% threshold; Rule R2 recommends that retailers with a medium-to-long-term contract share below 35% consider increasing this proportion; Rule R3 suggests that retailers with a high share of fixed-price customers should enhance load forecasting capabilities or optimize their retail tariff packages.

E. DISCUSSION

1) Advantages and Limitations of the Decision Tree Method

The experimental results validate the effectiveness of the decision tree algorithm in risk assessment for the electricity retail market. Compared to logistic regression, decision trees can capture nonlinear interaction effects among risk factors; for instance, the combined effect of the price spread ratio and load peak-to-valley ratio in Rule R1 far exceeds the linear sum of their individual impacts. Relative to support vector machines, decision trees maintain competitive prediction accuracy while providing transparent decision-making logic, which is particularly crucial when explaining risk assessment rationale to regulators and investors. Although a single decision tree achieves slightly lower prediction accuracy than random forests, it significantly reduces model complexity, shortening training time from approximately 43 seconds (on an Intel Core i7-12700K processor with 32 GB RAM and no GPU acceleration) to under 1 second, while markedly improving the simplicity of rule extraction.

The limitations of the decision tree method also warrant attention. Firstly, a single decision tree is highly sensitive to noise in the data; minor variations in the training set can lead to significant changes in the tree structure, a problem particularly pronounced with small sample sizes. Secondly, when dealing with high feature dimensions or strong correlations between features, the splitting preferences of decision trees may introduce bias in importance assessments. This paper mitigates these issues to some extent through domain knowledge filtering and correlation analysis during the feature engineering phase, but caution remains warranted in scenarios involving substantial increases in feature dimensions.

Regarding the robustness of decision tree splitting thresholds, this study employed the Bootstrap resampling method (1,000 repetitions) to estimate confidence intervals for key thresholds in the core rules. For rule R1, the 95% confidence interval for the spread ratio threshold at 18.7% is [17.2%, 20.1%], and that for the load peak-to-valley ratio threshold at 2.3 is [2.1, 2.5]; for rule R4, the 95% confidence interval for the spread ratio threshold at 12.3% is [11.1%, 13.6%]. The confidence interval widths for these thresholds range from 2 to 3 percentage points, indicating a certain degree of stability across different data subsets; however, in practical applications, dynamic adjustments should be made based on specific market conditions.

2) Implications for Risk Management in Electricity Retail Companies

From a practical perspective, the findings of this study offer several insights for risk management in electricity retail companies. Risk monitoring should focus on three core indicators – the price spread ratio, load peak-to-valley ratio, and the proportion of medium-to long-term contracts – rather than attempting to monitor all available metrics comprehensively. The formulation of risk management strategies should account for the distinct characteristics of different risk profiles: for price spread-driven risks, the emphasis should be on optimizing day-ahead bidding strategies and enhancing price forecasting; for load-driven risks, the focus should be on improving load prediction accuracy and optimizing customer mix; for contract structure-driven risks, the priority should be dynamic adjustment of power allocation ratios between medium-long term and spot markets. The allocation of risk management resources should follow a prioritization based on the support level and confidence level of risk profiles, giving priority to those with both high support and high confidence.

V. CONCLUSION

This study addresses the multi-source composite risks faced by electricity retailers in the retail market and proposes a risk assessment methodology based on the CART decision tree algorithm. The key findings are as follows: First, electricity retail market risks can be effectively characterized through 12 indicators covering wholesale, retail, and customer segments, with the day-ahead-to-real-time price spread ratio, load peak-to-valley ratio, and proportion of medium-to-long-term contracts serving as core risk determinants – these three factors collectively account for 46.9% of cumulative feature importance, providing robust data support for optimizing risk monitoring metrics. Second, the CART model achieves an accuracy of 87.8% $\pm$ 2.1% and an F1 score of 86.1% $\pm$ 2.3% in binary risk classification tasks, significantly outperforming logistic regression and support vector machines; compared to random forests, it substantially enhances model interpretability while maintaining limited accuracy loss. Repeated experiments and five-fold cross-validation (accuracy: 87.3% $\pm$ 2.0%; F1 score: 85.7% $\pm$ 1.9%) consistently demonstrate strong model robustness, indicating CART decision trees achieve an optimal balance between predictive accuracy and decision transparency. Third, explicit risk criteria derived from the decision tree can be formulated in "if-then" structure. Rules R1–R4 delineate typical risk scenarios – including price-difference-load dual compression, inadequate contract protection, and prediction bias combined with retail demand rigidity – providing tailored risk mitigation strategies for retailers with varying risk preferences. For example, under Rule R1, electricity retailers may proactively increase their holdings of medium-and long-term contracts when the price spread approaches 18.7% and the load peak-to-valley ratio exceeds 2.3, thereby locking in part of the costs; under Rule R4, when the price spread falls below 12.3% and the load curve becomes more flat, they may appropriately raise their participation in the spot market to achieve higher returns. Bootstrap resampling validation results demonstrate that the threshold values for these core rules exhibit good stability.

Future research can be conducted in the following directions: At the data level, higher-frequency real-time transaction data, as well as textual unstructured information such as policy announcements and news reports, can be incorporated to further diversify risk signal sources; at the algorithmic level, hybrid model architectures combining decision trees with deep learning can be explored, utilizing deep learning networks for feature extraction and decision trees for interpretable classification; at the application level, the proposed risk assessment framework can be deployed as an online early warning system and its timeliness and robustness can be validated in real-world market environments.

FOUNDATION ITEMS

This work was supported by Science and Technology Project of State Grid Sichuan Electric Power Company: 521996240003.

REFERENCES

[1] N. Sinha, C. Lucheroni, and C. Mari, "BME Model: Forecasting Electricity Supply and Demand Curves Using Disentangled Prices and Volumes," Applied Stochastic Models in Business and Industry, vol. 41, no. 4, Art. no. e70032, 2025, doi: https://doi.org/10.1002/ASMB.70032.

[2] S. Zheng, J. Wang, and Y. Gong, "Research on Automatic Identification of Power Big Data Anomalies Based on Improved Decision Tree," International Journal of High Speed Electronics and Systems, vol. 34, no. 4, 2025, doi: https://doi.org/10.1142/S0129156425403341.

[3] J. Jafferi, O. Hiromitsu, and S. A. Razalli, "Real-Time Electricity Retail Pricing Dual Optimization With Context-Based Fuzzy Optimal Algorithm," Electric Power Components and Systems, vol. 50, no. 6–7, pp. 374–385, 2022, doi: https://doi.org/10.1080/15325008.2022.2136785.

[4] M. Nizami, M. Hossain, R. B. Amin, et al., "A residential energy management system with bi-level optimization-based bidding strategy for day-ahead bi-directional electricity trading," Applied Energy, vol. 261, Art. no. 114322, 2020, doi: https://doi.org/10.1016/j.apenergy.2019.114322.

[5] B. Chen, T. Fu, L. Wei, et al., "Data-Driven Risk Warning of Electricity Sales Companies in the Whole Business Process," Sustainability, vol. 17, no. 9, Art. no. 3884, 2025, doi: https://doi.org/10.3390/SU17093884.

[6] Z. Jiang, O. Ozcelebi, E. R. Khoury, et al., "Identifying uncertainty factors that affect U.S. retail electricity sales dynamics," Financial Innovation, vol. 12, no. 1, Art. no. 133, 2026, doi: https://doi.org/10.1186/S40854-026-00948-3.

[7] F. Wenjie, S. Hongtao, B. Bo, et al., "Management of Power Marketing Audit Work Based on Tobit Model and Big Data Technology," Mobile Information Systems, vol. 2022, 2022, doi: https://doi.org/10.1155/2022/1375331.

[8] J. Liu, "Risk warning for customer request tickets in the power marketing system based on user profile analysis," Office Automation, vol. 31, no. 10, pp. 16–18, 2026, doi: https://doi.org/10.3969/j.issn.1007-001X.2026.10.006.

[9] X. Jia, P. Ji, F. Chen, et al., "A distributionally robust bilevel optimization model for wholesale-retail electricity market design," Scientific Reports, vol. 16, no. 1, Art. no. 1768, 2026, doi: https://doi.org/10.1038/S41598-025-29971-7.

[10] N. Sinha and C. Lucheroni, "Demand and supply curve forecasting using a monotonic autoencoder for short-term day-ahead electricity market bid curves," Applied Energy, vol. 397, Art. no. 126262, 2025, doi: https://doi.org/10.1016/J.APENERGY.2025.126262.

[11] C. Zhao, Study on Power Retailers' Electricity Purchase-Sale Transactions and Vertical Cooperation-Based Emission Reduction Strategies Based on Game Theory. Wuhan University, 2022, doi: https://doi.org/10.27379/d.cnki.gwhdu.2022.003785.

[12] F. Hu, Y. Zhao, Y. Yu, et al., "Deep reinforcement learning-based strategic bidding in electricity markets via variational autoencoder-assisted competitor behavior learning," Engineering Applications of Artificial Intelligence, vol. 161, Art. no. 112205, 2025, doi: https://doi.org/10.1016/J.ENGAPPAI.2025.112205.

[13] Y. Yu, Y. Gao, Y. Li, et al., "Interpretable data-driven contingency classification for real-time corrective security-constrained economic dispatch," IET Renewable Power Generation, vol. 18, no. 3, pp. 476–488, 2024, doi: https://doi.org/10.1049/rpg2.12830.

[14] Z. Zhou, P. Xu, H. Zhang, et al., "Extraction and Interpretability of Reinforcement Learning Control Strategies for Distribution Network Fault Recovery Based on Decision Tree," in 2025 IEEE 14th Data Driven Control and Learning Systems, 2025, pp. 1179–1184.

[15] L. Luo, Y. Song, Y. Liu, et al., "The research on intelligent safety risk assessment and early warning mechanisms for power systems based on big data and artificial intelligence technology," Advances in Resources Research, vol. 5, no. 2, pp. 666–688, 2025, doi: https://doi.org/10.50908/ARR.5.2_666.

[16] V. Albani, P. Moreira, M. Geremia, et al., "A hybrid stochastic-machine learning model for long-term predictions of electricity prices," Renewable Energy Focus, vol. 58, Art. no. 100832, 2026, doi: https://doi.org/10.1016/J.REF.2026.100832.

[17] O. Wenhua, C. Yaqi, L. Qiong, et al., "Multi-Scenario Clustering-Based Distribution Network with Distributed Power Sources Operational Risk Warning Method," Journal of Physics: Conference Series, vol. 2557, no. 1, 2023, doi: https://doi.org/10.1088/1742-6596/2557/1/012045.