Construction of A Big Data-Driven Risk Identification, Assessment, and Precise Control System for Electricity Bill Accounting

Songyan Du, Xionghui Zheng, Xinling Zheng, Xingye Lin, Xupeng Chen, Yan Lin
State Grid Fujian Marketing Service Center (Metering Center), Fuzhou 350001, Fujian, China
Corresponding author: Xionghui Zheng (e-mail: yxfwzx614@163.com)

ABSTRACT Ensuring the security of electricity revenue and improving operational efficiency are critical issues in modern power systems, particularly under increasingly complex electromagnetic energy transmission environments and data-intensive grid operations. To address the lagging response and passive management characteristics of traditional electricity bill accounting risk control, this study proposes a big data-driven "identification–evaluation–control" framework for electricity bill accounting. By integrating multi-source operational data, machine learning techniques, and process reengineering strategies, the proposed framework transforms conventional accounting management into a proactive and high-precision risk control system. The study clarifies the core architecture, technical support mechanisms, and implementation pathway of the system, emphasizing data fusion, risk identification, quantitative assessment, precise intervention, and dynamic optimization. Furthermore, the framework enables accurate prediction, scientific evaluation, and efficient control of electricity bill accounting risks through continuous data-driven decision-making. The proposed approach provides a practical solution for power enterprises facing increasingly diversified settlement mechanisms and operational uncertainties, while offering methodological references for intelligent risk management and reliable operation in data-centric power and energy systems.

INDEX TERMS big data, electricity bill accounting risk, multi-source data fusion, risk assessment, electromagnetic energy systems

I. INTRODUCTION

A. RESEARCH BACKGROUND

As the lifeblood of the national economy, electricity revenue provides the essential cash flow for power enterprises to support the high-load energy demands of the manufacturing sector, ensuring that stable power supplies and grid investments directly underpin the industry's continuous production and sustainable development [1]. With the continuous promotion of the power market reform, the types of power users are increasingly diversified (including residents, industry and commerce, new energy enterprises, etc.), and the electricity bill settlement mode is constantly innovated (such as time-sharing price, peak valley price, market-oriented transaction price, etc.). In addition, the explosive growth of power data in the context of the energy Internet has significantly enhanced the complexity, concealment and conductivity of electricity bill accounting risks [2]. According to the "2023 China Power Industry Risk Management Report", the average electricity bill recovery rate of Chinese power enterprises is 97.5%, with a bad debt rate of about 2.3%. Only large-scale power enterprises have annual bad debt losses exceeding 10 billion yuan; The frequent occurrence of risk events such as user payment delays, contract breaches, fraudulent electricity use, and policy adjustments has become a core bottleneck restricting the improvement of revenue quality for power companies.

B. RESEARCH SIGNIFICANCE

1) Theoretical Significance

Constructing a five in one theoretical framework for electricity billing risk control, which includes data fusion, risk identification, quantitative evaluation, precise control, and dynamic optimization, filling the gap in the systematic research of big data risk control in the power industry and enriching the theoretical connotation of power marketing risk management.

Deepen the research on the application of multi-source data fusion and machine learning algorithms in power accounting risk control, reveal the inherent mechanism of big data-driven risk prevention and control, and improve the research paradigm and method system of electricity billing management.

Establish a four-dimensional evaluation model of "risk type impact degree occurrence probability controllability", break through the limitations of traditional qualitative evaluation, and provide new theoretical references for quantitative risk assessment in the power industry.

2) Practical Significance

Accurately addressing industry pain points such as difficulty in recovering electricity bills, lagging risk warning, and hidden fraudulent behavior, we have improved the electricity bill recovery rate to over 99% and reduced the bad debt rate by over 40% through a full process risk control mechanism, ensuring that the core revenue of power enterprises is "stored in the warehouse".

Build a standardized and replicable big data risk control operation system, reduce the labor and collection costs of power companies' risk control (expected to decrease by more than 30%), and improve risk control efficiency and accuracy. To provide practical solutions for power enterprises to cope with new challenges such as diversified users and complex settlement brought about by market-oriented reforms, to help enterprises stabilize their operating cash flow, and to provide financial support for power grid construction and energy supply guarantee. By aligning high-precision energy data with the specific load characteristics of production, this approach shifts power industry risk control from an "experience-driven" model to a "data-driven" paradigm, offering a scalable blueprint for the digital and intelligent upgrading of revenue management across energy-intensive industrial sectors.

Combined with the characteristics of income management in the power industry, expressions such as "income recovery guarantee" and "full and timely payment" are adopted, which not only comply with the financial industry's expression standards for income security but also fit the practical scenarios of electricity bill accounting and recovery in the industry, improving the professionalism and accuracy of expression.

C. TECHNICAL ROADMAP

The technical route of this article is as follows:

1. Literature research stage: Sort out the research results in related fields such as electricity accounting risks, big data risk control, and power marketing, and clarify the research status and shortcomings.
2. Theoretical construction stage: Integrating multidisciplinary theories, constructing a five in one risk control system framework, and revealing risk control mechanisms.
3. System design phase: Design data fusion framework, risk identification and assessment model, and full process control strategy.
4. Empirical research stage: Select typical power enterprises, collect data for empirical research, and verify the effectiveness of the system.
5. Strategy optimization stage: Based on empirical results, optimize the risk control system and control strategies.
6. Conclusion and Prospect Stage: Summarize the research results, point out the research limitations and future research directions.

II. THEORETICAL BASIS AND ANALYTICAL FRAMEWORK

A. DEFINITION OF CORE CONCEPTS

1) Electricity Billing Risks

Electricity bill accounting risk refers to the possibility that electricity bills cannot be collected in full and on time, and revenue may be damaged due to various factors such as users, enterprises, policies, and markets throughout the entire process of electricity bill accounting, collection, settlement, and recovery. Its core features include objectivity, concealment, conductivity, and controllability [3,4].

2) Big Data-Driven Risk Control for Electricity Bill Accounting

Big data-driven risk control in electricity bill accounting refers to the use of big data technology to integrate multiple sources of data such as electricity marketing, user credit, government information, and social behavior. Through algorithms such as data mining and machine learning, accurate identification, scientific evaluation, and efficient management of electricity bill accounting risks are achieved, ensuring full and timely recovery of electricity bills. Its core goal is to enhance the comprehensiveness, accuracy, real-time performance, and foresight of risk control.

3) Multi Source Data Fusion

Multi source data fusion refers to the collection, cleaning, conversion, and correlation analysis of heterogeneous data such as internal marketing data of power enterprises and external government affairs, credit, social behavior, etc., to form a unified and high-quality risk control data resource, providing comprehensive data support for risk identification and assessment.

4) Precise Control

Precise control refers to the development of differentiated and targeted control measures based on risk identification and assessment results, targeting different risk types, levels, and user groups, to achieve refined risk control of "one set of plans for one type of risk, one strategy for one user".

B. THEORETICAL BASIS

1) Risk Management Theory

Risk management theory emphasizes the effective control of risks and the minimization of losses through risk identification, assessment, control, monitoring, and other processes. The risk control of electricity bill accounting driven by big data follows the risk management logic of "identification evaluation control optimization", and enhances the scientificity and efficiency of each link through technological empowerment [5- 7].

2) Data Fusion Theory

The theory of data fusion aims to integrate heterogeneous data from multiple sources, eliminate data redundancy and conflicts, and improve data quality and information value. Multi source data fusion provides a data foundation for comprehensive identification of electricity bill accounting risks, and can explore associated risks that are difficult to detect with a single data.

3) Machine Learning Theory

Machine learning theory uses algorithms to enable computers to learn patterns from data, achieving automated prediction and decision-making. The risk recognition model based on machine learning can automatically capture risk features in data, improving the accuracy and efficiency of recognition [8].

4) Customer Relationship Management Theory

The theory of customer relationship management emphasizes putting customers at the center and improving customer satisfaction and loyalty through differentiated services and management. The design of precise control strategies follows this theory, and differentiated measures are developed for different risk levels and user types to maintain customer relationships while preventing and controlling risks.

C. CONSTRUCTION OF FIVE IN ONE RISK CONTROL SYSTEM FRAMEWORK

Build a five in one framework for electricity bill accounting risk control system consisting of "data fusion risk identification quantitative evaluation precise control dynamic optimization". The core connotations and interrelationships of each dimension are as follows:

1) Dimension of Data Fusion

Data fusion is the foundation of the risk control system, which includes three stages: data collection, preprocessing, and fusion

1. Data collection: Integrate internal marketing data and external multi-source data of power enterprises to build comprehensive risk control data resources [9];
2. Data preprocessing: Clean, deduplicate, complete, and standardize the collected data to improve data quality;
3. Data fusion: By using methods such as correlation analysis and feature extraction, heterogeneous data is fused into a unified set of risk control features [10-12].

2) Risk Identification Dimensions

Risk identification is the core of the risk control system, based on the fusion of data and machine learning algorithms, to achieve precise positioning of risk types and risk users:

1. Feature engineering: Extract multidimensional risk features such as user payment behavior, electricity usage characteristics, and credit status;
2. Model construction: Build an ensemble learning model to identify risks such as payment delays, fraudulent electricity usage, and contract breaches;
3. Real time monitoring: Establish a real-time data monitoring mechanism to capture risk signals in a timely manner [13].

Real-time monitoring adopts a hierarchical mechanism of "core data T+0 real-time monitoring + auxiliary data quasi-real-time update":

Core data (internal data): Key risk-related data such as user electricity consumption behavior, payment records, and settlement data are collected and monitored in real-time (T+0) to ensure the timely capture of core risk signals (e.g., abnormal electricity consumption, overdue payments);

Auxiliary data (external data): Non-immediately changing data such as credit data and policy data are updated with an optimized frequency ("credit data T+1 update + policy data T+3 update"), and a data update trigger mechanism (e.g., real-time synchronization during policy adjustments) is established to avoid lag effects.

3) Quantitative Evaluation Dimensions

Quantitative assessment is the key to the risk control system, which achieves the quantitative classification of risk levels through multidimensional indicators and scientific methods

1. Indicator system: Construct a four-dimensional evaluation index of "risk type impact degree occurrence probability controllability";
2. Evaluation method: Using combination weighting method and fuzzy comprehensive evaluation method to quantify the risk level;
3. Classification: Classify risks into four levels: low, medium, high, and extremely high, providing a basis for control strategies [14-15].

4) Precise Control Dimension

Precise control is the goal of the risk control system, and based on the evaluation results, formulate full process control measures:

1. Pre warning: give early warning to high-risk users and take preventive measures;
2. In process interception: Real time interception of risky behaviors during the electricity bill settlement process;
3. Post event disposal: Take measures such as collection and punishment for risk events that have already occurred.

5) Dynamic Optimization Dimension

Dynamic optimization is the guarantee of risk control system, continuously optimizing system performance through data feedback:

1. Data update: Real time update of multi-source data to ensure data timeliness;
2. Model iteration: Based on new data and risk changes, iteratively optimize the identification and evaluation model;
3. Strategy adjustment: dynamically adjust control measures based on control effectiveness [16].

The interrelationships of five dimensions: data fusion provides fundamental support for risk identification; Risk identification is a prerequisite for quantitative assessment; Quantitative evaluation provides a basis for precise control; The effect of precise control is achieved through dynamic optimization feedback, continuously improving system performance, and forming a closed-loop logic of "data identification evaluation control optimization".

D. ANALYSIS OF TYPES AND CAUSES OF ACCOUNTING RISKS IN ELECTRICITY CHARGES

1) Core Risk Types

Through the analysis of the electricity bill accounting data of a provincial power company from 2020 to 2023, combined with industry practice, four core risk types have been identified:

1. Payment delay risk (accounting for 62.3%): Due to difficulties in fund turnover, weak payment awareness, and inconvenient payment channels, users fail to pay electricity bills on time, resulting in delayed electricity bill collection and occupying enterprise funds;
2. Fraudulent electricity consumption risk (accounting for 18.7%): Users maliciously evade paying electricity bills through methods such as electricity theft, tampering with electricity meters, and false reporting of electricity consumption types, causing direct economic losses to the enterprise;
3. Contract breach risk (accounting for 12.5%): Users violate the provisions of the power supply and consumption contract, such as changing the nature of electricity consumption without authorization, exceeding the capacity of electricity consumption, etc., resulting in disputes and difficulties in electricity fee settlement and recovery;
4. Policy and market risks (accounting for 6.5%): Due to external factors such as adjustments in electricity pricing policies, changes in market-oriented trading rules, and macroeconomic fluctuations, users' payment ability may decrease or settlement disputes may arise [17].

2) Analysis of Risk Causes

1. At the user level: difficulties in capital turnover, weak awareness of payment, lack of integrity, and deviation in understanding policies;
2. At the enterprise level: data fragmentation, untimely risk identification, single control measures, and inconvenient payment channels;
3. External environmental factors: macroeconomic downturn, adjustment of electricity pricing policies, deepening of market-oriented reforms, imperfect social credit system, etc;
4. Technical aspects: Traditional risk control technology is outdated, data processing capabilities are insufficient, and there is a lack of real-time monitoring methods [18-19].

3) Advantages of Big Data Technology Adaptability

1. Advantages of data integration: Breaking down data barriers, integrating internal and external multi-source data, and achieving comprehensive risk identification;
2. Advantages of precise identification: Capturing risk characteristics through machine learning algorithms to enhance the ability to identify hidden and associated risks;
3. Real time monitoring advantages: realizing real-time data collection and dynamic analysis, early warning of risks, and improving response speed;
4. Advantages of scientific evaluation: Multi dimensional quantification of risk levels, providing scientific basis for the formulation of control measures;
5. Advantages of precise control: Based on user profiles and risk levels, develop differentiated control strategies to improve control efficiency.

III. DESIGN OF MULTI SOURCE DATA FUSION FRAMEWORK

A. DATA SOURCES AND CLASSIFICATION

Build a multi-source data resource library of "internal data+external data", covering six major categories of data, as follows:

Internal data of power marketing (core data)

1. User basic information: user name, type (residential/commercial/new energy), electricity address, contact information, power supply and consumption contract information, etc;
2. Electricity consumption behavior data: electricity consumption, load curve, voltage and current data, distribution of electricity consumption periods, information on electrical equipment, etc;
3. Accounting settlement data: electricity bill amount, payment history, payment channels, arrears records, late fee payment status, etc;
4. Service interaction data: repair records, complaint records, consultation records, business processing records, etc. [20].

1) External Associated Data (Supplementary Data)

1. Social credit data: enterprise credit reports, personal credit scores, information on dishonest persons subject to enforcement, administrative penalty records, etc;
2. Government open data: business registration information, tax registration information, social security payment information, industry access qualifications, etc;
3. Economic operation data: macroeconomic indicators, industry prosperity index, regional economic development data, enterprise revenue data, etc;
4. Social behavior data: user consumption behavior data, social media data, travel data, etc;
5. Policy environment data: electricity price adjustment policies, market-oriented trading rules, environmental protection policies, industrial support policies, etc. [21].

B. DATA COLLECTION AND PREPROCESSING

1) Data Collection Plan

1. Internal data collection: Real-time collection of marketing, accounting, and service data through internal platforms

(power marketing system, electricity consumption information collection system, customer service system) with a frequency of T+0 (real-time) - T+1 (quasi-real-time);

2. External data collection: Batch collection of external data through government data sharing platforms, third-party credit institution interfaces, and public data crawling, with a frequency of T+1 (credit data) - T+3 (policy data), dynamically adjusted according to data update frequency.3. Data collection standards: Develop unified data collection standards, clarify data formats, field definitions, and collection frequencies to ensure data consistency.

2) Data Preprocessing Process

1. Data cleaning: Adopt methods such as missing value filling (mean/median/machine learning prediction), outlier removal (box plot method, Z-score method), and duplicate value deletion to address data quality issues;

2. Data standardization: Min Max standardization and Z-score standardization are used for numerical data, and single hot encoding and label encoding are used for categorical data to ensure that the data is compatible with the model input;

3. Data integration: integrate data from different sources and in different formats into a unified database, establish a unique user ID (such as user number, ID number/unified social credit code), and realize data association;

4. Data desensitization: desensitize user privacy data (such as ID number and contact information) to ensure data security and compliance [22].

C. DATA FUSION AND FEATURE ENGINEERING

1) Data Fusion Methods

Adopting a two-level fusion architecture of "feature layer fusion+decision layer fusion":

1. Feature layer fusion: Extract features and perform correlation analysis on preprocessed multi-source data to construct a user 360 degree portrait feature set, including user basic features, electricity consumption behavior features, accounting features, credit features, social behavior features, etc;

2. Decision level fusion: Weighted fusion of risk identification results from different data sources to enhance the reliability of risk identification [23].

The risk identification model adopts a fusion architecture of "real-time core features + quasi-real-time auxiliary features": Core features first drive preliminary risk judgment, and after auxiliary data is updated, secondary verification and risk level correction are performed, ensuring both real-time response speed and evaluation accuracy. For example, when a user's abnormal electricity consumption is monitored (T+0 core data), a warning is triggered first, and the risk level is accurately determined after supplementing T+1 updated credit data.

2) Feature Engineering Design

1. Basic features: user type, electricity usage years, contract period, electricity capacity, etc;

2. Characteristics of electricity consumption behavior: daily average electricity consumption, load fluctuation coefficient, peak valley electricity consumption ratio, number of sudden changes in electricity consumption, deviation rate from historical electricity consumption during the same period, etc;

3. Accounting features: historical payment punctuality rate, number of outstanding payments, maximum outstanding duration, outstanding amount, preferred payment channels, etc;

4. Credit characteristics: credit score, number of dishonest records, number of administrative penalties, abnormal business records, etc;

5. Social behavior characteristics: consumption ability level, business activity, number of complaints and reports, industry prosperity correlation, etc;

6. Derivative features: the cross characteristics of electricity consumption behavior and accounting features (such as high electricity consumption+low payment punctuality rate), time series features (such as the trend of payment delay times in the past 3 months), etc. [24-25].

The final construction of a risk control feature set containing 128 dimensional features, including 76 numerical features and 52 categorical features, provides comprehensive support for the risk identification model.

D. DATA STORAGE AND SECURITY ASSURANCE

1) Data Storage Architecture

Adopting a hybrid architecture of "distributed storage+relational database":

1. Distributed storage (Hadoop HDFS): stores massive amounts of unstructured data (such as electricity consumption curve data, user behavior logs) and semi-structured data (such as policy documents, credit reports);

2. Relational database (MySQL/PostgreSQL): stores structured data (such as user basic information, accounting settlement data) to ensure the efficiency of data query and correlation analysis;

3. Data Warehouse (Hive): Integrate various types of data, build a risk control themed data mart, and support multi-dimensional analysis.

2) Data Security Assurance

1. Permission management: Establish a hierarchical authorization mechanism, clarify the data access permissions of different roles, and prevent data leakage;

2. Data encryption: Encrypt transmission and storage data (using SSL/TLS protocol for transmission encryption and AES algorithm for storage encryption);

3. Security audit: Record data access and operation logs, conduct regular security audits, and promptly detect abnormal operations;

4. Compliance guarantee: Adhere to laws and regulations such as the Data Security Law and the Personal Information Protection Law, and standardize the process of data collection and use [26-28].

IV. CONSTRUCTION OF RISK IDENTIFICATION AND ASSESSMENT MODEL DRIVEN BY BIG DATA

A. RISK IDENTIFICATION MODEL DESIGN

Build a risk identification model with "basic model+integrated optimization", integrate multiple machine learning algorithms, and improve recognition accuracy and generalization ability.

1) Basic Model Selection

Select four classic machine learning algorithms as the basic models and fully leverage the advantages of each algorithm:

1. Logistic regression (LR): The model is simple, highly interpretable, and suitable for capturing linear risk features;
2. Random Forest (RF): Strong anti overfitting ability, suitable for handling high-dimensional features and nonlinear relationships;
3. Gradient Boosting Tree (XGBoost): High prediction accuracy, able to effectively capture feature interaction effects;
4. Support Vector Machine (SVM): It performs well in small sample data and is suitable for handling complex classification problems.

2) Integrated Model Construction

Using the "Stacking Integration" strategy to construct the final risk identification model, the process is as follows:

1. Data partitioning: Divide the dataset into training and testing sets in a 7:3 ratio, and use 5-fold cross validation to optimize model parameters;
2. Basic model training: Train four types of basic models separately on the training set and output probability prediction results;
3. Meta model training: Train a logistic regression meta model using the predicted results of the base model as new features, and output the final risk identification results;
4. Model optimization: Using a combination of grid search and Bayesian optimization methods to optimize the hyperparameters of each model and improve model performance [29-30].

3) Model Training and Validation

1. Training data: Select 1 million user data from a provincial power company from 2020 to 2023, including 150000 high-risk user samples and 850000 normal user samples;
2. Evaluation indicators: Accuracy, Precision, Recall, F1 Score, and Area Under the ROC Curve (AUC) are used as model evaluation indicators;
3. Verification results: The accuracy of the integrated model reached 96.8%, the precision was 95.7%, and the recall was 94.3%, F1-Score95.0%,AUC0.982, Significantly better than a single baseline model (accuracy 85% -92%).

The model design fully considers the strong regulatory requirements of the power industry. In addition to outputting identification results, it synchronously generates a "feature contribution report" to provide clear decision-making basis for control measures such as power restriction and debt collection, ensuring the transparency and compliance of control behaviors.

B. CONSTRUCTION OF HIERARCHICAL RISK ASSESSMENT SYSTEM

1) Design of Evaluation Indicator System

Construct a four-dimensional evaluation index system of "risk type impact degree occurrence probability controllability", covering 16 secondary indicators, as follows:

2) Determination of Indicator Weights

Adopting a combination weighting method combining Analytic Hierarchy Process (AHP) and Entropy Weight Method, balancing subjective experience and objective data:

1. Subjective Weighting (AHP): Invite 10 experts from fields such as power marketing, risk management, and finance to construct a judgment matrix and calculate subjective weights;
2. Objective weighting (entropy weighting method): Based on historical data, calculate the information entropy and weight of each indicator;
3. Combination weight: Combine subjective and objective weights in a 4:6 ratio to obtain the final indicator weight.

The combination weight ratio of 4:6 (subjective:AHP / objective:Entropy Weight Method) is based on the risk assessment principle of "data-driven as the main approach, with experience judgment as a supplement":

Objective weight (Entropy Weight Method) accounts for 60%: The core of electricity bill accounting risk assessment relies on historical data (e.g., user payment records, electricity consumption behavior data, default records). The Entropy Weight Method can objectively reflect the differentiation and importance of indicators through data information entropy, avoiding human experience bias, which is consistent with the core positioning of big data-driven risk control;

Subjective weight (AHP) accounts for 40%: For risk factors such as policy changes and market fluctuations that are difficult to quantify, adjustments need to be made based on industry expert experience to compensate for the limitations of purely data-driven evaluation, ensuring the evaluation results are both objective and practical.

3) Classification of Risk Levels

Using the fuzzy comprehensive evaluation method to calculate the comprehensive risk score (0-100 points), and dividing the risk into four levels:

1. Low risk (0-30 points): The probability of risk occurrence is extremely low, the impact is small, and the controllability is strong;

TABLE 1. Electricity Bill Accounting Risk Assessment Index System

Primary Indicator	Secondary Indicator	Indicator Description	Indicator Type	Weight
Risk Type	Payment Delay Risk	Risk of users failing to pay electricity bills on time	Qualitative + Quantitative	0.35
Risk Type	Electricity Fraud Risk	Risk of users deliberately evading electricity bills	Qualitative + Quantitative	0.30
Risk Type	Contract Breach Risk	Risk of users violating power supply and consumption contracts	Qualitative + Quantitative	0.20
Risk Type	Policy and Market Risk	Risk caused by changes in external policies and market conditions	Qualitative	0.15
Impact Degree	Economic Loss Scale	Amount of electricity bill loss caused by risk occurrence	Quantitative	0.40
Impact Degree	Impact Scope	Number of users and regional scope affected by the risk	Quantitative	0.30
Impact Degree	Transmission Effect	Chain reaction of the risk on other links	Qualitative	0.30
Occurrence Probability	Historical Occurrence Frequency	Number of historical occurrences of similar risks	Quantitative	0.45
Occurrence Probability	Risk Trigger Intensity	Strength of factors leading to risk occurrence	Qualitative + Quantitative	0.35
Occurrence Probability	Trend Change	Changing trend of risk occurrence probability	Quantitative	0.20
Controllability	Effectiveness of Control Measures	Control effect of existing measures on the risk	Qualitative	0.40
Controllability	Response Speed	Response and disposal efficiency after risk occurrence	Quantitative	0.30
Controllability	Recovery Cost	Recovery cost after risk disposal	Quantitative	0.30

2. Medium risk (30-60 points): The probability of risk occurrence is moderate, the impact range is limited, and the controllability is strong;

3. High risk (60-85 points): The probability of risk occurrence is high, the economic losses are significant, and the controllability is average;

4. Extremely high risk (85-100 points): The probability of risk occurrence is high, the losses are severe, the transmission effect is strong, and the controllability is weak.

4) Risk Identification and Assessment Process

1. Data input: Input the fused feature data into the risk recognition model;

2. Risk identification: The model outputs a list of risk users and risk types;

3. Indicator scoring: Based on the evaluation indicator system, calculate the scores of each indicator;

4. Comprehensive evaluation: Using the fuzzy comprehensive evaluation method to calculate the compre hensive risk score and determine the risk level;

5. Result output: Generate a risk identification and assessment report, including risk user information, risk types, risk levels, core triggers, etc.

V. DESIGN OF PRECISE CONTROL STRATEGY FOR THE ENTIRE PROCESS

Based on the full process risk control logic of "pre warning, in-process interception, and post disposal", differentiated and precise control strategies are developed for different types and levels of risks, forming a closed-loop risk control mechanism.

A. PRE WARNING: RISK PREVENTION AND SOURCE CONTROL

1) Risk Warning Mechanism

1. Warning threshold setting: Based on the risk assessment results, set warning thresholds for different types of risks (such as the risk warning threshold for payment delay being "2 or more payment delays in the past 3 months+credit score below 600 points");

2. Warning triggering method: Real time monitoring of user data, automatic triggering of warnings when the warning threshold is met, and generation of warning information;

3. Alert information push: Push alert information to risk control specialists and customer managers through system messages, SMS, emails, and other methods.

Source control measures

1. Customer access control: For newly applied electricity users, risk assessment is conducted based on multi-source data, and high-risk users are required to provide guarantees or prepay security deposits;

2. Refined contract management: Develop differentiated contract terms for users with different risk levels (such as shortening payment cycles for high-risk users);

3. Payment guidance: push personalized payment reminders to users (based on payment habits and preferences), promote convenient payment channels (such as automatic withholding and online payment);

4. Credit cultivation: Conduct credit education for low - and medium risk users, promote payment obligations and the consequences of dishonesty.

Optimization of payment channels (targeting enterprise-level cause: inconvenient payment channels):

Expand diversified payment channels: Add 8 types of channels including third-party payment, automatic deduction, and offline convenience points, covering different user groups (e.g., reserving cash payment windows for elderly users);

Personalized channel recommendation: Push adaptive payment methods based on user behavior data (e.g., preference for online operation/offline processing) to improve convenience;

Real-time response to channel issues: Establish a fast response mechanism for payment channel failures (response time ≤ 2 hours) to avoid payment delays caused by system problems.

B. IN PROCESS INTERCEPTION: REAL-TIME MONITORING AND DYNAMIC INTERVENTION

Real time monitoring mechanism

1. Accounting settlement monitoring: Real time monitoring of the electricity bill settlement process, identifying abnormal settlement behavior (such as frequent changes to payment accounts, large outstanding fees not paid);

2. Electricity consumption behavior monitoring: Real time analysis of user electricity consumption curves to identify fraudulent electricity consumption characteristics

(such as sudden changes in electricity consumption and abnormal load curves);

3. Abnormal behavior warning: Real time triggering of warnings and initiation of intervention processes for abnormal behavior detected by monitoring.

Dynamic intervention measures

1. Payment restrictions: For high-risk users, non counter payment channels are restricted, and they are required to pay in person at the business hall and provide relevant proof;

2. Electricity restrictions: For users with extremely high risks of fraudulent electricity use, temporary mea sures such as power restrictions and outages will be taken to prevent violations;

3. Real time communication: Communicate with users in real-time about abnormal situations through phone calls, text messages, and other means to verify payment willingness and ability;

4. Policy interpretation: Promptly interpret policies to users and negotiate payment plans for risks caused by policy adjustments.

C. POST DISPOSAL: RISK RESOLUTION AND LONG-TERM CONTROL

1) Graded Collection Strategy

Develop differentiated collection plans for different risk levels and user types:

1. Low risk users: SMS reminders, automatic voice collection, collection frequency is 3 days and 7 days after overdue fees;

2. Medium risk users: Customer managers call for collection and conduct on-site visits, with collection frequencies of 1 day, 3 days, and 7 days after overdue payments;

3. High risk users: lawyer's letter collection, cooperation with third-party collection agencies, collection frequency is initiated within 1 day after arrears;

4. High risk users: apply for payment orders, file lawsuits, be included in the list of dishonest users, and recover according to law.

Risk mitigation measures

1. Payment scheme negotiation: For users who have difficulties, negotiate and formulate installment payment and deferred payment schemes, and sign repayment agreements;

2. Debt restructuring: For users with large outstanding debts, risks can be resolved through debt transfer, guarantee exchange, and other means;

3. Policy assistance: For users affected by policies and market shocks, connect with relevant support policies to assist users in overcoming difficulties;

4. Fraudulent disposal: For fraudulent electricity users, the overdue fees shall be recovered, fines shall be imposed, and those with serious circumstances shall be transferred to judicial authorities.

Long term control measures

1. Risk file establishment: Establish specialized files for risk users to record the entire process of risk occurrence, disposal, and resolution;

2. Credit linkage: Linking user payment behavior with the social credit system, allowing trustworthy users to enjoy payment discounts and green channels for business processing, and implementing joint punishment for untrustworthy users;

3. Regular review: Regularly review risk events, analyze their causes and disposal effects, and optimize the risk control system;

4. Continuous monitoring: For high-risk users who have been disposed of, continuously monitor their payment and electricity usage behavior to prevent risk recurrence.

D. DYNAMIC OPTIMIZATION MECHANISM

1. Data update: Real time update of internal and external data to ensure data timeliness and comprehensiveness;

2. Model iteration: Based on new data and risk changes every quarter, iteratively optimize the risk identification and evaluation model to improve performance;

3. Strategy adjustment: dynamically adjust control measures and parameters (such as warning thresholds and collection frequency) based on control effectiveness and market changes;

4. Effect evaluation: Establish a risk control effect evaluation index system (such as risk identification accuracy, bad debt rate, electricity fee recovery rate), regularly evaluate and optimize it.

A dynamic adjustment mechanism for risk weights is established, with quarterly reviews and optimizations based on risk occurrence frequency, impact scope, and market changes. When market-oriented reforms deepen or policy adjustments are frequent, the mechanism can automatically increase the weight of policy and market risks to ensure the adaptability of the evaluation system.

VI. EMPIRICAL RESEARCH

A. EMPIRICAL DESIGN

1) Research Object

Selecting a provincial power company as the empirical research object, the company serves over 20 million households, covering various types of users such as residents, industry and commerce, and new energy. The total electricity revenue in 2023 exceeds 80 billion yuan, which is a typical representative.

2) Data Sources and Processing

1. Data scope: From 2020 to 2023, the company's internal data such as user basic information, electricity consumption behavior, and accounting settlement, as well as external data obtained from government platforms and credit institutions, cover a total of 1 million user samples;

2. Data processing: According to the preprocessing process in section 3.2, complete data cleaning, standardization, and fusion, and construct a 128 Vitesse collection;

3. Experimental grouping: Use 2020-2022 data as the training set (700000 samples), and 2023 data as the testing set (300000 samples); Compare the risk control effects

between 2022 (traditional risk control mode) and 2023 (big data risk control system).

3) Evaluation Indicator System

Constructing a four-dimensional evaluation index system of "recognition effect evaluation effect control effect comprehensive effect":

TABLE 2. Empirical Evaluation Index System

Primary Indicator	Secondary Indicator	Indicator Description	Calculation Method
Identification Effect	Risk Identification Accuracy (%)	Quantitative Indicator	(Number of correctly identified high-risk users / Total identified high-risk users) × 100%
Identification Effect	Risk Identification Recall Rate (%)	Quantitative Indicator	(Number of correctly identified high-risk users / Actual number of high-risk users) × 100%
Identification Effect	Early Warning Lead Time (hours)	Quantitative Indicator	Warning time before risk occurrence
Assessment Effect	Risk Level Classification Accuracy (%)	Quantitative Indicator	(Number of users with matching assessed level and actual risk / Total number of users) × 100%
Assessment Effect	Consistency of Assessment Results (%)	Quantitative Indicator	(Number of users with consistent results in different assessment cycles / Total number of users) × 100%
Control Effect	Electricity Bill Recovery Rate (%)	Quantitative Indicator	(Actually recovered electricity bills / Receivable electricity bills) × 100%
Control Effect	Bad Debt Rate (%)	Quantitative Indicator	(Bad debt amount / Receivable electricity bills) × 100%
Control Effect	Collection Cost Reduction Rate (%)	Quantitative Indicator	(Traditional collection cost - Big data collection cost) / Traditional collection cost × 100%
Comprehensive Effect	Comprehensive Score	Quantitative Indicator	Comprehensive effect score calculated based on weighted indicators

B. EMPIRICAL RESULTS AND ANALYSIS

1) Comparison of Recognition Effects

TABLE 3. Comparison of Risk Identification Effects

Indicator	Traditional Risk Control Model (2022)	Big Data-Driven Risk Control System (2023)	Improvement Range
Risk Identification Accuracy (%)	82.3	96.8	+14.5 percentage points
Risk Identification Recall Rate (%)	78.5	94.3	+15.8 percentage points
Early Warning Lead Time (hours)	24	96	+72 hours

According to Table 3, the recognition effect of the big data risk control system is significantly better than that of the traditional mode:

1. The accuracy of risk identification has increased by 14.5 percentage points, reaching 96.8%, effectively reducing the misjudgment rate and unnecessary intervention on normal users;
2. The recall rate for risk identification has increased by 15.8 percentage points, reaching 94.3%, significantly improving the ability to identify hidden and associated risks;
3. The advance warning time has been increased from 24 hours to 96 hours, providing sufficient time for the implementation of risk control measures.

The average warning response time for core risks of the system is 2 hours, which is 85% shorter than the traditional mode (13.3 hours), verifying the effectiveness of the layered real-time mechanism.

2) Evaluation Effect Comparison

TABLE 4. Comparison of Risk Identification Effects

Indicator	Traditional Risk Control Model (2022)	Big Data-Driven Risk Control System (2023)	Improvement Range
Risk Level Classification Accuracy (%)	75.6	92.5	+16.9 percentage points
Consistency of Assessment Results (%)	72.3	90.8	+18.5 percentage points

According to Table 4, the evaluation effect of the big data risk control system has significantly improved:

1. The accuracy of risk level classification has increased by 16.9 percentage points, reaching 92.5%, providing a scientific basis for differentiated control strategies;
2. The consistency of the evaluation results has been improved by 18.5 percentage points, reaching 90.8%, ensuring the stability and reliability of the evaluation results.

3) Comparison of Control Effects

TABLE 5. Comparison of Risk Control Effects

Indicator	Traditional Risk Control Model (2022)	Big Data-Driven Risk Control System (2023)	Improvement/Reduction Range
Electricity Bill Recovery Rate (%)	95.1	99.2	+4.1 percentage points
Bad Debt Rate (%)	2.3	1.35	-0.95 percentage points (41.3% reduction)
Collection Cost Reduction Rate (%)	-	38.7	+38.7 percentage points

According to Table 5, the control effect of the big data risk management system is significant:

1. The electricity fee recovery rate has increased from 95.1% to 99.2%, an increase of 4.1 percentage points, significantly increasing the revenue of power companies;
2. The bad debt ratio decreased from 2.3% to 1.35%, a decrease of 41.3%, significantly reducing economic losses;
3. The collection cost has been reduced by 38.7%, and ineffective manpower and time investment have been reduced through precise collection.

After adjusting the weight of policy and market risks, the system's identification accuracy for such risks increased from 82.1% to 93.4%, and the bad debt rate caused by related risks decreased by 52.3%, proving that the adjusted weight is more in line with actual risk management needs.

4) Comprehensive Effect Comparison

TABLE 6. Comparison of Comprehensive Effects (Unit: Points, 100-Point Scale)

Primary Indicator	Traditional Risk Control Model (2022)	Big Data-Driven Risk Control System (2023)	Improvement Range
Identification Effect	76.8	95.2	+18.4 points
Assessment Effect	73.5	91.6	+18.1 points
Control Effect	78.2	94.8	+16.6 points
Comprehensive Score	76.2	93.9	+17.7 points

According to Table 6, the comprehensive score of the big data risk control system is 93.9 points, which is 17.7 points higher than the traditional risk control mode's 76.2 points. The comprehensive effect is significantly better than the traditional mode, verifying the effectiveness and practicality of the big data-driven risk control system.

5) Empirical Conclusion

1. The big data-driven electricity billing risk control system can significantly improve risk control effectiveness, achieving comprehensive breakthroughs in recognition accuracy, scientific evaluation, and precise control;

2. Multi source data fusion and machine learning models are the core support for improving risk control effectiveness, which can effectively capture hidden risks and associated risks;

3. The closed-loop control strategy throughout the entire process can achieve pre risk prevention, in-process intervention, and post risk resolution, significantly reducing bad debt rates and improving electricity recovery rates;

4. This system has strong industry applicability and can provide replicable solutions for risk management of electricity bill accounting for other power enterprises.

VII. CONCLUSION AND PROSPECT

By deeply integrating big data technology with the power consumption profiles of manufacturing, this article constructs a "five-in-one" risk control system comprising "data fusion, risk identification, quantitative evaluation, precise control, and dynamic optimization" to safeguard electricity bill accounting. This framework leverages real-time production data from enterprises to transition from traditional manual oversight to an intelligent, automated revenue protection model. Through theoretical analysis, system design, and empirical research, the following core conclusions are drawn:

The core types of electricity billing risks include payment delay risk, fraudulent electricity use risk, contract breach risk, and policy and market risk. The causes involve four levels: users, enterprises, external environment, and technology. Traditional risk control models are difficult to adapt to complex risk prevention and control needs.

2. A risk identification model based on Stacking integration strategy has an accuracy of 96.8% and a recall rate of 94.3%, significantly better than a single machine learning model; The constructed "four-dimensional integrated" evaluation system achieves scientific and quantitative classification of risk levels, with an accuracy rate of 92.5% in level classification.

The proposed "pre warning, in-process interception, and post disposal" full process control strategy, through differentiated and precise measures, has increased the electricity recovery rate to 99.2%, reduced the bad debt rate by 41.3%, and lowered the collection cost by 38.7%. The comprehensive effect is significantly better than traditional risk control models.

Future research can be conducted in the following directions:

Expand the scope of empirical research, select power enterprises of different regions and scales to conduct research and verify the universality of the system;

Optimize the risk identification model, incorporate force majeure factors, and enhance the model's antiinterference ability;

Expand the dimension of data fusion, integrate user industry chain, supply chain and other data, and explore deeper risk characteristics;

Conduct input-output analysis to quantify the economic benefits of the big data risk control system; Combining emerging technologies such as artificial intelligence and blockchain to further enhance the intelli gence and security of the risk control system; This paper explores prevention and control strategies for emerging electricity bill accounting risks, such as market-oriented transaction settlement risks, to address the evolving power procurement needs of manufacturing under electricity marketization reforms. By aligning risk management with the industry's specific energy consumption patterns, this approach ensures that financial controls adapt to the shifting regulatory and operational demands.

AUTHOR CONTRIBUTIONS

Conceptualization –Songyan Du, Xionghui Zheng, Xinling Zheng, Xingye Lin, Xupeng Chen and Yan Lin; methodology – Songyan Du, Xionghui Zheng, Xinling Zheng, Xupeng Chen and Yan Lin; investigation – Songyan Du, Xionghui Zheng, Xinling Zheng, Xingye Lin, Xupeng Chen and Yan Lin; writing-original draft preparation – Songyan Du, Xionghui Zheng, Xinling Zheng, Xingye Lin, Xupeng Chen and Yan Lin. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. Mazzucco and D. Dyachuk, "Balancing electricity bill and performance in server farms with setup costs," Future Generation Computer Systems, vol. 28, no. 2, pp. 415-426, 2011, doi: 10.1016/j.future.2011.04.015.

[2] V. Reddy K S S and M. Kumari S, "Prediction-Based Optimal Sizing of Battery Energy Storage Systems in PV Integrated Microgrids for Electricity Bill Minimization," Journal of The Institution of Engineers (India): Series B, vol. 103, no. 5, pp. 1733-1745, 2022, doi: 10.1007/s40031-022-00778-8.

[3] Y. Liu, H. Hou, K. Zhou, et al., "Research on Construction of Management and Control System of Electricity Bill Risks in Power Grid Enterprises under Energy Conservation and Emission Reduction Principle," International Journal of Business and Management, vol. 6, no. 8, pp. 287, 2011, doi: 10.5539/ijbm.v6n8p287.

[4] M.S, "J, H.R.E.H," B, Huadong M, et al. Optimization of electric spring operational strategy to minimize electricity bill. Electric Power Systems Research, pp. 201, 2021, doi: 10.1016/j.epsr.2021.107540 21130.

[5] Loria F , Brinca P , Iskrev N I .On Identification Issues in Business Cycle Accounting Models[J].SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5119365.

[6] A.E, "N, Alok V K," Cloud Automation: New Era of Electricity Bill Automation using GPRS and Web Interface. International Journal of Computer Applications, vol. 29, no. 11, pp. 38-41, 2011, doi: 10.5120/3686-5111.

[7] Kumar K R , Shreyaa R R , Thiyaneswaran K M ,et al.Smart Run Time Electricity Bill Calculation using LabVIEW[J].International Journal of Innovative Technology and Exploring Engineering, 2020, doi: 10.35940/ijitee.d1855.039520.

[8] T. Gerpott J and I. Mahmudova, "Determinants of price mark-up tolerance for green electricity - lessons for environmental marketing strategies from a study of residential electricity customers in Germany," Business Strategy and the Environment, vol. 19, no. 5, pp. 304-318, 2010, doi: 10.1002/bse.646.

[9] W. Zhao, "Application of Principal Component Analysis Algorithm in Abnormal Diagnosis of Electricity Bill Reading and Receiving Data," in State Grid Henan Electric Power Company Marketing Service Center. 2024, doi: 10.4108/ eai.8-12-2023.2344734.

[10] M. Fikru G, "Estimated electricity bill savings for residential solar photovoltaic system owners: Are they accurate enough? Applied Energy," 2019; 253:113501, doi: 10.1016/j.apenergy.2019.113501.

[11] M. Fikru G, "Electricity bill savings and the role of energy efficiency improvements: A case study of residential solar adopters in the USA," Renewable and Sustainable Energy Reviews, vol. 106, pp. 124-132, 2019, doi: 10.1016/ j.rser.2019.02.028.

[12] Y. Wang, X. Wang, and Y. Zhang, "Leveraging thermal storage to cut the electricity bill for datacenter cooling," in University of Tennessee, Knoxville, University of Tennessee, Knoxville and The Ohio State University, University of Tennessee, Knoxville. 2011, doi: 10.1145/2039252.2039260.

[13] Reddy M P , Gorain R , Ganguly G .Drive Cycle Type Identification Using Data Driven Methodologies[J].SAE International, 2026, doi: 10.4271/2026-26-0649.

[14] Chen J , Cai Y .A Study of Big Data Prediction in Accounting Management Risks[J].Applied Mathematics and Nonlinear Sciences, 2024, 9(1), doi: 10.2478/amns-2024-0913.

[15] M. Weng and Weng D .Discuss the Accounting Information Risks and Preventive Measures Based on Big Data[J], "2021,", doi: 10.2991/assehr.k.210728.022.

[16] D. Eryilmaz and S. Gafford, "Can a daily electricity bill unlock energy efficiency? Evidence from Texas," The Electricity Journal, vol. 31, no. 3, pp. 7-11, 2018, doi: 10.1016/j.tej.2018.03.009.

[17] A.M. Khatib, I. M. Mahable, S.S. Maniyar, et al., "IoT based Automatic Electricity bill generation with theft detection," Journal of Trend in Scientific Research and Development, vol. 2, no. 3, pp. 2586-2589, 2018, doi: 10.31142/ijtsrd11416 21131.

[18] S. José B H, M. Ascensión M C G, and S. David H, "Clear and transparent design proposal for the regulated electricity bill in Spain," grafica, vol. 7, no. 13, pp. 11-11, 2018, doi: 10.5565/rev/grafica.125.

[19] E. @ A F N, A. Pauzi M, H. Yusri M, et al., "Disaggregated Electricity Bill Base on Utilization factor and Time-of-use (ToU) Tariff," International Journal of Electrical and Computer Engineering (IJECE), vol. 7, no. 3, pp. 1498-1505, 2017, doi: 10.11591/ ijece.v7i3.pp1498-1505.

[20] Cui H .A DBSCAN Anomaly Data Identification Method Driven by Data Distribution[J].Applied and Computational Engineering, 2025, 134(1):73-79, doi: 10.54254/2755-2721/2025.20861.

[21] K. Muralitharan, R. Sakthivel, and Y. Shi, "Multiobjective optimization technique for demand side management with load balancing approach in smart grid," Neurocomputing, vol. 177, pp. 110-119, 2016, doi: 10.1016/j.neucom.2015.11.015.

[22] S. Huh, J. Woo, S. Lim, et al., "What do customers want from improved residential electricity services? Evidence from a choice experiment," Energy Policy, vol. 85, pp. 410-420, 2015, doi: 10.1016/j.enpol.2015.04.029.

[23] J. Octavio R B, G. Miguel J L, B. Emilio E, et al., "IoT-Based Electricity Bill for Domestic Applications," Sensors (Basel, Switzerland), vol. 20, no. 21, pp. 6178, 2020, doi: 10.3390/s20216178.

[24] S. Vipin and K. Sudhir S, "Modeling and Design of Data Cube for Electricity Bill Deposit System," International Journal of Computer Applications, vol. 80, no. 11, pp. 11-14, 2013, doi: 10.5120/13904-1736.

[25] C. Sibiya A, K. Kusakana, and B. Numbi P, "TRU Energy Monitoring for Potential Cost Saving in Electricity Bills for Cathodic Protection Units: South African Case," International Journal of Electrical and Electronic Engineering & Telecommunications, vol. 10, no. 2, pp. 145-150, 2021, doi: 10.18178/ijeetc.10.2.145-150.

[26] A. Zaman U and A. Mahbub A O, "Modeling Electricity Bill with the Reflection of CO2 Emissions and Methods of Implementing AMI for Smart Grid in Bangladesh," International Journal of Intelligent Systems and Applications (IJISA), vol. 14, no. 5, pp. 14-21, 2022, doi: 10.5815/ijisa.2022.05.02.

[27] J. Miota C, S. Khadem, and M. Bahloul, "Deep reinforcement learning based electricity bill minimization strategy for residential prosumer," Mathematics and Computers in Simulation, vol. 238, pp. 296-305, 2025, doi: 10.1016/ j.matcom.2025.06.011.

[28] K. Li, N. Zhou, L. Wang, et al., "A Customer-Focused Electricity Bill Pay Points Siting Model," Advanced Materials Research, vol. 588-589, pp. 2064-2068, 2012, doi: 10.4028/www.scientific.net/AMR.588-589.2064.

[29] C. Cárdenas C, E. Gaytán A A, and T. Ozuna, "The crucial role of electricity bill comprehension in consumer intention of solar panel systems adoption," Energy Policy, vol. 210, Art. no. 115064, 2026, doi: 10.1016/j.enpol.2025.115064 21132.

[30] V. Kumar A, E, and N. A, "Secure Electricity Bill Automation Using GPRS and Web Interface in Conjunction with Elliptic Curve Crypto-Stegano Scheme," International Journal of Computer Theory and Engineering, vol. 4, pp. 510-513, 2012, doi: 10.7763/ijcte.2012.v4.521 21133.