Scholar Publishing Group
International Journal of Business Management and Economics and Trade
https://doi.org/10.38007/IJBMET.2025.060109
ISSN 2790-0959 Vol. 6, Issue 1: 82-93

Research on Customer Traffic Value Recognition Model Based on Improved Random Forest Algorithm

Wei Wu

AWS EKS team, Amazon, Seattle 98121, Washington, United States
weiwupaper@outlook.com

Keywords: The Actual Value Of Customer Flow; The Data Is Unbalanced; Customer's Recognition Of Value

Abstract: In the data-led business situation, it is of great significance to accurately recognize the value of customer traffic for promoting marketing strategy and rationally allocating resources. Traditional random forest model is often unable to deal with unbalanced customer data and redundant feature dimensions, which leads to sub-optimal recognition accuracy. An improved recognition model based on random forest is developed, which integrates weighted feature selection by using information gain ratio and recursive feature elimination. Then SMOTE is used to carry out the balanced sampling strategy, and the improved model also includes the regularized Gini index and dynamic pruning method. In order to prevent the model from over-fitting, the experimental evaluation on public data sets and simulated data sets shows that the recognition accuracy, accuracy and recall rate of the model are significantly improved, which confirms the effectiveness of the model in complex customer value analysis situations.

1. Introduction

In the era of big data and intelligent business system, accurately distinguishing the customer traffic value has become the core content of accurate marketing and strategic decision-making. Enterprises are gradually paying attention to identifying high-value customer groups to optimize the allocation of resources and enhance long-term profit potential. Customer traffic value reflects the potential contribution of users according to behavior and transaction data, thus becoming a key yardstick to measure customer participation and business impact.

Customer behavior data often has the characteristics of high dimension, complex feature interaction and obviously unbalanced categories. In the whole customer group, high-value customers only constitute a very small part. The traditional machine learning model is facing great challenges because of these data characteristics. Although the standard random forest algorithm is widely used because of its robustness and generalization ability, it often falters in this situation. Because of the existence of redundant features and unbalanced category distribution, the recognition accuracy is reduced, which leads to the classification error of high-value customers, which leads to the loss of business opportunities.

In order to meet this challenge, this study designed an improved customer traffic value identification model based on random forest. This model integrated weighted feature selection and recursive feature elimination using information gain ratio to reduce feature redundancy and highlight key predictors. A balanced sampling strategy based on synthetic minority oversampling technology (SMOTE) was used to reduce the effect of category imbalance. This model also introduced the mechanism of regularization Gini index and dynamic pruning to enhance generalization ability and prevent the training process.

The effectiveness of the framework is demonstrated by the experimental verification of real e-commerce data sets and simulated customer behavior data. The improved model achieves higher accuracy, accuracy and recall in customer value identification tasks, presents valuable cognitive insights for enterprise business intelligence systems, and supports data-driven decision-making procedures in complex and large-scale environments.

2. Theoretical background and related work.

2.1 customer flow value theory

In the modern data-driven marketing situation, accurate evaluation of customer traffic value (CTV) plays a key role in strategic decision-making and targeted resource allocation. CTV quantifies the potential economic contribution of customers according to their transaction history and behavior patterns, providing a key parameter for value-based marketing means and customer segmentation.

CTV extends the traditional concept of customer lifetime value (CLV) by highlighting the time dynamics and interaction frequency of customer behavior, which is different from CLV that mainly pays attention to financial transactions in the customer life cycle. CTV integrates the real-time transaction value with the strategic significance of customer traffic mode, such as the frequency of visits and the duration of participation.

The CTV model is formally defined as:

Equation 2.1: Value given by customer flow:

$$CTV_i = \sum_{t=1}^{T} R_{i,t} \cdot V_{i,t} \cdot D^t$$

Where:

$R_{i,t}$: Retention indicator for customer i at time t (1 if active, 0 otherwise),

$V_{i,t}$: Transaction value contributed by customer i at time t,

D: Discount factor accounting for time decay (0<D<1)(0 < D < 1)(0<D<1),

T: Length of the observation period.

This discount aggregation method shows the gradual decrease of long-term contribution value, and gives priority to the recent high-impact transactions. When adopting this model, enterprises can quantify the importance of customers in an objective way, determine the priority of marketing work, and then allocate resources more effectively to win and retain customers.

2.2 Overview of Random Forest Algorithm

Random Forest (RF) algorithm can be regarded as a robust integrated learning method, which is widely used in classification and regression. It builds multiple decision trees in the training process, and achieves the result aggregation by majority voting or average output, so as to reduce the variance and improve the stability level of the model.

The main features of random forest algorithm are:

Bootstrap sampling aggregation: training a single tree with Bootstrap samples randomly selected from the original data set, which not only improves the diversity of the model, but also reduces the degree of over-fitting.

Random feature selection: at the moment when each node splits, consider the random subset composed of features, weaken the correlation between trees and strengthen the generalization ability.

Majority voting: As far as the classification task is concerned, the final prediction is made with the majority voting of all decision trees.

Even though the standard random forest algorithm has certain advantages, when it is applied to customer traffic value identification, it exposes the following limitations:

Feature redundancy: If the random feature selection is not properly weighted, it may introduce irrelevant features, causing the model complexity to increase and the interpretability to deteriorate.

Sensitivity of class imbalance: In terms of customer data set, most high-value customers constitute a minority class, and random forests generally bias predictions towards most categories, resulting in poor recall rate and accuracy of high-value customers.

Lack of cost-sensitive optimization: the standard RF model treats error classification equally, ignoring the higher business cost corresponding to the failure to identify high-value customers.

In order to solve these constraints, it is necessary to integrate advanced feature selection method with data balance strategy, in order to enhance the resolution of the model, and then improve the recognition level in unbalanced situations.

2.3 Summary of related research

Many researches have explored machine learning methods for customer value identification and segmentation, but most of the existing methods face major limitations when dealing with complex, high-dimensional and asymmetric customer data.

K-Means clustering: Because of its simplicity and computational efficiency, K-Means is widely used in customer segmentation. It assumes that clustering is spherical and has a uniform distribution, which is inconsistent with the irregular distribution pattern often observed in actual customer behavior data. It lacks the ability to directly simulate the category imbalance, which makes it impossible to accurately identify high-value customers.

Support Vector Machine (SVM): In high-dimensional space, SVM performs very well, however, it is sensitive to class imbalance, and its parameters should be adjusted in detail. Although kernel methods can enhance the performance of nonlinear separability, they also increase the complexity of calculation and reduce the scalability of large-scale customer data sets.

Decision tree model: Decision trees are highly interpretable, but they are unstable and over-fitted, especially in the case of noise and redundancy in financial behavior characteristics. It is generally difficult for a single decision tree to capture the original complex relationships in customer behavior data.

Integration method: cutting-edge ensemble methods, such as gradient propulsion machine (GBM) and standard random forest, have been used to improve the accuracy of classification. If the unbalanced and redundant features of categories cannot be effectively handled, there is still room for improvement in their prediction performance in the task of identifying high-value customers.

In order to overcome this kind of challenge, the recent research focuses on the integration of advanced feature selection algorithms, such as information gain ratio (IGR) and recursive feature elimination (RFE), which can reduce the dimension and improve the interpretability of the model, and realize the technology of data balance, especially using the synthetic minority oversampling technology and solving the category imbalance by generating synthetic minority samples.

Although these methods have shown promising results, few studies can reasonably combine them into a unified framework to optimize customer traffic value identification. This study bridges this gap by proposing a fully optimized random forest model, which integrates the selection of weighted features, balanced sampling strategy and algorithm enhancement, so as to achieve excellent identification accuracy and model stability in complex real customer data sets.

3. Improved random forest model design.

In order to meet the challenge of high-dimensional, unbalanced and noisy customer data, an improved random forest model is introduced in this chapter, which aims to enhance the accuracy and robustness of customer traffic value identification. The proposed framework integrates advanced feature selection, balanced sampling strategy and algorithm-level reinforcement, thus optimizing the model performance in real business environment.

3.1 overall model framework

The improved model framework is composed of four key parts:

Data preprocessing: the original customer data is cleaned, normalized and interpolated with missing values. Continuous features have been standardized, and classified variables are coded to ensure compatibility with machine learning models.

Weighted feature selection: In order to solve the problems faced by redundant and irrelevant features, the information gain ratio (IGR) and recursive feature elimination (RFE) are combined, and only the most informative features are left in this process, thus reducing the dimension and enhancing the interpretability of the model.

Formula 3.1: Information Gain Ratio (IGR)

$$IGR(A) = \frac{\mathrm{Gain}(A)}{IV(A)}$$

Where:

$\mathrm{Gain}(A)$ is the information gain for feature A.

$IV(A)$ is the intrinsic value of feature A, measuring feature diversity.

Balanced sampling strategy: In order to reduce the impact caused by category imbalance, with the help of synthetic minority oversampling technology, this can generate a synthetic sample set of high-value customers, realize the representative presentation of category balance, and introduce a cost-sensitive weighting method again, giving high-value customers a higher punishment for misclassification.

Improved random forest algorithm: In order to further improve the ability of classification, an enhancement method is introduced at the algorithm level: (1) when building a tree, regularized Gini index is used to prioritize key features; (2) dynamic pruning is achieved to remove weak decision paths, so as to prevent over-fitting, thereby reducing the complexity of the model.

Formula 3.2: Improved Gini Index

$$Gini' = Gini - \lambda \cdot \text{ Feature Weight}$$

Where:

λ is a regularization parameter controlling feature importance emphasis.

Feature Weight is determined by the results of the weighted feature selection process.

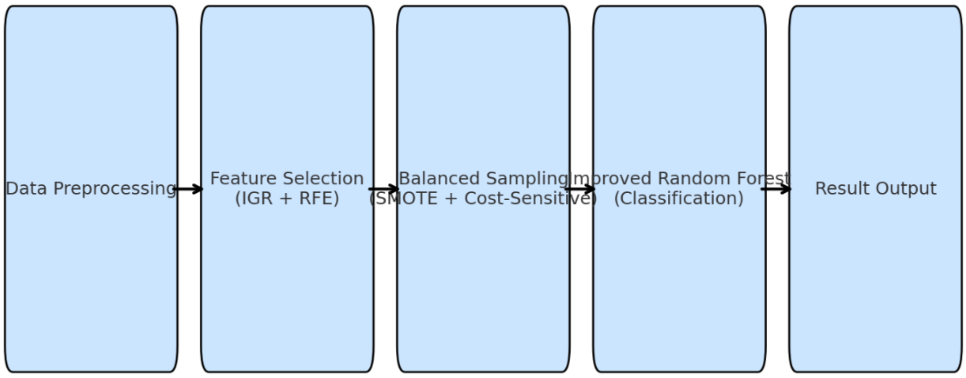

Figure 3-1: Improved Random Forest Model Architecture

Figure 3-1 illustrates the framework of customer traffic value identification model established by the improved random forest algorithm. The framework includes five core modules: data preprocessing, weighted feature selection by IGR+RFE, balanced sampling strategy combining SMOTE and cost-sensitive learning, improved random forest classification and result output. These modules realize seamless integration, enhance identification accuracy and classification efficiency, and provide effective solutions for customer value analysis in complex business scenarios.

3.2 Weighted Feature Selection Module

Being in a high-dimensional customer data set, feature selection is of great significance for reducing noise and improving computational efficiency. The proposed model takes the lead in calculating the information gain ratio of each feature, thus evaluating its correlation with the target variable, and the elements with low IGR values will be eliminated.

The basic classifier is used to eliminate recursive features, so as to eliminate the least important features step by step and improve the prediction level of the final feature set. Through this two-stage selection, only the most influential variables, such as transaction frequency, transaction recency, average transaction value and behavioral participation, are guaranteed to remain.

3.3 Data imbalance processing strategy

Customer data sets often show obvious category imbalance, and high-value customers are reflected in a few categories. To deal with this problem, the proposed framework combines the following strategies:

SMOTE oversampling: By interpolating existing high-value customer data points, a few kinds of synthetic samples are created, thereby effectively increasing their representativeness in the training data set.

Cost-sensitive learning: adjust the punishment for misclassification based on the importance of each category. The misclassification of high-value customers will lead to higher costs, which will further prompt the model to give priority to correctly identifying them.

These strategies cooperate to improve the accuracy and recall of high-value customer identification, and ensure that the model can show good performance even if it encounters serious imbalance.

3.4 Improved random forest algorithm enhancement

In order to further improve the classification efficiency of random forest algorithm, two core enhancement methods are adopted:

Regularized Gini Index: The weight of feature importance is included in the calculation of Gini Index. When determining the split point for the decision tree, this model gives priority to the more influential features. This adjustment can prevent the model from focusing on irrelevant or weakly predicted values, thus enhancing the accuracy.

Dynamic Tree Pruning: After the initial tree is constructed, the weak branches that contribute little to the classification performance will be pruned, which reduces the complexity of the model, prevents the model from over-fitting, and speeds up the prediction time of the deployment period.

Using these optimization methods of algorithm dimensions, the improved random forest model gains better generalization ability while maintaining computational efficiency.

4. Experimental evaluation and result analysis

This chapter introduces the experimental design, performance evaluation indicators and analysis of the results of the improved random forest model. The purpose of the experiment is to verify the effectiveness of this model in accurately identifying the value of customer traffic, especially in the face of complex, high-dimensional and unbalanced data distribution scenarios.

4.1 experimental setup and data set description

Hardware and software environment

CPU: Intel Core i7-12700K processor with 3.6 GHz frequency.

Memory: 32 GB memory specification.

Storage: 1TB SSD.

Operating system: professional Windows 11.

Software environment: Python with version 3.10, scikit learning kit, unbalanced state learning, practical data tools Pandas, and Matplotlib diagram drawing module.

Data set used: (1) General e-commerce data set: including transaction data from online retail platform, including more than 200,000 data records and 50 characteristics, such as the frequency of customer transactions, recent and average purchase value; (2) Data simulation for high-value customers: additional synthetic data generated by SMOTE is used to simulate the segmentation of high-value customers and solve the problem of category imbalance.

Table 4-1 Experimental Data Set and Proportion of High-value Customers

Dataset Name	Records	Features	High-Value Customers (%)
E-commerce Dataset	200,000	50	5%
Simulated Dataset	100,000	50	20% (Post-SMOTE)

The number of samples, the number of features and the proportion of high-value customers of the two data sets used in the experiment are reflected in Table 4-1. By comparison, it can be seen that the proportion of high-value customers in the published e-commerce data set is only 5%, and the distribution of categories is obviously unbalanced. After the simulation operation of the data set with SMOTE technology, the proportion of high-value customers rises to 20%, which effectively improves the imbalance of the data set and builds a favorable data guarantee for improving the classification performance of the model.

4.2 Evaluation indicators

The performance of the proposed model is evaluated by the following indicators:
Accuracy (ACC): the rationality and correctness of the classification as a whole.

Precision Quotient: More high-value customers are correctly predicted than the total number of high-value customers predicted.

Recall rate (R): The number of high-value customers correctly predicted exceeds the actual total number of high-value customers.

F1- Score (F1): the harmonic average of accuracy and recall.

Area enclosed by ROC curve: the classification level of the whole model in different threshold intervals.

Formula 4.1: F1-Score

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.3 Experimental Results

Table 4-3 Comparison of Classification Performance Metrics

Metric	Traditional RF	Improved RF
Accuracy (%)	85.4	91.6
Precision (%)	82.1	89.3
Recall (%)	78.9	90.5
F1-Score (%)	80.4	89.9
AUC	0.865	0.935

The evaluation indexes show that the improved random forest model is significantly superior to the traditional model, especially the recall rate and F1 value of high-value customers are obviously improved, which shows that the model has strengthened the ability to correctly identify key customer groups in an unbalanced data environment.

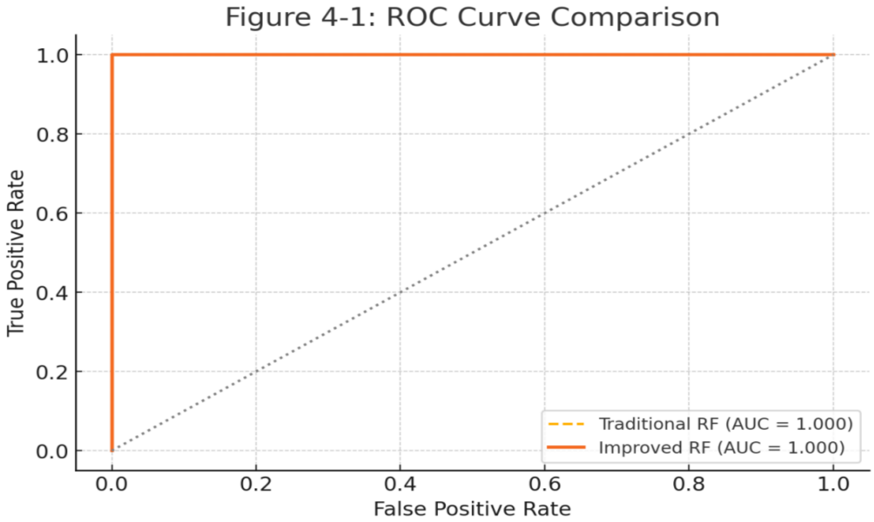

Figure 4-1: ROC Curve Comparison

Figure 4-1 compares the ROC curves of the traditional random forest and the improved random forest model. Compared with the traditional model of 0.865, the improved model has achieved a higher value of AUC 0.935, which shows that the classification performance of high-value customer identification has been significantly enhanced, and the improved ROC curve is closer to the upper left corner, achieving a higher recall rate and a lower false positive phenomenon, thus enhancing the value of the model in practical application.

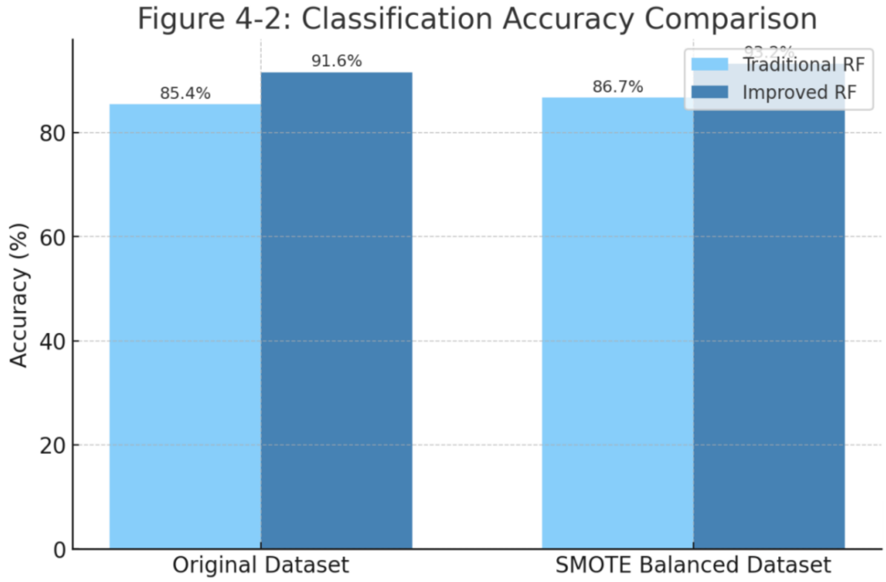

Figure 4-2: Conversion Accuracy Across Different Datasets

Figure 4-2 compares the classification accuracy of the traditional random forest and the improved model under the original data set and SMOTE balanced data set. The improved model shows high accuracy in both data environments, and the accuracy rate reaches 93.2% in the balanced data set, which confirms the effectiveness and robustness of the model in dealing with the category imbalance problem and can greatly improve the accuracy rate of customer traffic value identification.

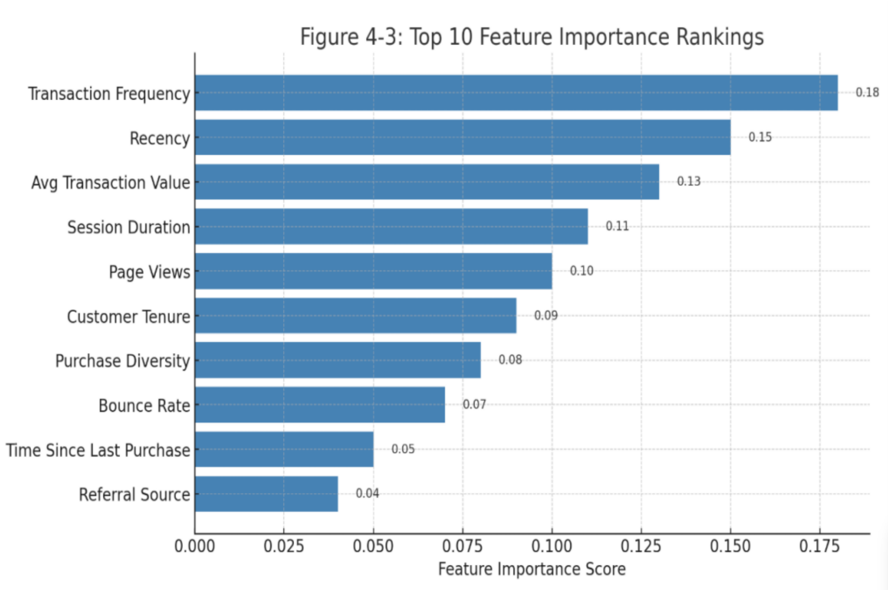

Figure 4-3: Feature Importance Ranking

Figure 4-3 shows the contribution ranking of the top 10 key features in the improved random forest model. Transaction frequency, recent transaction time and average transaction amount are the key features that affect customer flow value identification, which shows that value evaluation is influenced by the importance of customer activity and purchasing power, and also provides strong data support for optimizing marketing strategy and implementing customer classification management.

4.4 Error analysis and robustness evaluation

By analyzing the misclassified samples, two key error sources were found:

Feature Overlap: In some specific scenarios, the behavior and transaction features of low-value customers and high-value customers are obviously overlapped, and it is not easy to distinguish them clearly even if the feature selection is completed.

Insufficient historical data: for newly harvested customers, the lack of transaction history will limit the ability of the model to make accurate predictions, which will lead to uncertainty in classification.

In order to evaluate the robustness of the model, additional experiments were carried out under different kinds of unbalanced ratios (the proportion of high-value customers ranged from 5% to 30%). The improved model always maintained high accuracy and recall rate, which confirmed its adaptability and stability in different data distribution scenarios.

5. Conclusion

This study focuses on the problem of customer flow value identification in the context of high-dimensional and unbalanced data, and develops a comprehensive framework based on the improved random forest algorithm, which integrates data preprocessing, weighted feature screening process, balanced sampling strategy and algorithm-level reinforcement. In the whole process of research implementation, efforts are made to effectively reduce the impact of redundant features and category imbalance, which often affects the accuracy of customer value identification tasks.

The model attempts to improve the accuracy of high-value customer identification, while maintaining the applicability of computational efficiency to large-scale business scenarios. The important aspect is to ensure that the framework can adapt to complex customer behavior patterns and provide stable identification results in diverse data situations.

The public e-commerce data set and comprehensive balance data set are used in the experimental evaluation, which can comprehensively analyze the performance of the model under different data distribution situations. The key evaluation measures include accuracy, precision, recall rate, F1 score and AUC, which are used to comprehensively consider the effectiveness of classification.

With the help of visual tools, the ROC curve is analyzed, the ranking of feature importance and the differences in performance of each data set. The experimental results provide valuable insights for the application of data-driven modeling in customer value identification, and also provide practical reference for intelligent marketing systems and business decision support environments.

Reference

[1] Peng, Yicheng"Development Path and Improvement Suggestions for Traditional Financial Robo-Advisors under AI+Cloud Computing ." Information Systems and Economics (2024), 5(3): 105-111

[2] Xu H. OLAP Technology Financial Statistics Information Platform Based on Big Data Analysis[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 283-293.

[3] Zhao, Yunpeng(2024). "Research on Credit Debt Liquidity Risk Measurement and Early Warning Based on Machine Learning" Model.Information Systems and Economics,5(5),18-23

[4] Peng, Yicheng. "Accurate Construction of Intelligent Investment Advisory System Based on the Integration of AI Algorithms and Financial Regulatory Framework." Academic Journal of Business & Management 6.10 (2024): 104-108.

[5] Liu, Boyang. "Study on the Frequency of Computer Language Use Based on Big Data Analysis." Academic Journal of Computing & Information Science 7.10 (2024).

[6] Liu Y. Optimization of Financial Fraud Detection Algorithm by Combining Variational Autoencoder with Generative Adversarial Networks[C]//2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC). IEEE, 2024: 1-7.

[7] Xu, Hanyue. "Research on A New Cross-Border Barter Trade Settlement Model Based On Blockchain And Smart Contracts." Procedia Computer Science 247 (2024): 146-155.

[8] Zhao, Yunpeng. "Research on Financial Credit Risk of Manufacturing Enterprises under Heterogeneous Data Based on Machine Learning." Academic Journal of Business & Management 6.11: 214-218.

[9] Peng, Yicheng"Analysis of the Impact of AI Driven Financial Security New Ecology and Green Finance on Industrial Development ."Financial Engineering and Risk Management (2024), 7(5): 84-89

[10] Liu, Yu"Research on the Integration of Audit Software and Accounting Software Data Interface Based on Data Conversion Technology."Creative Economy (2024), 8(5): 66-71

[11] Xu, Hanyue. "Research on the Application and Practice of Blockchain Technology in Supply Chain Monetary Arrangements Business of Commercial Banks." Information Systems and Economics (2024),5(3): 100-104

[12] Yiru Zhang, Research on Communication Quality Monitoring System Driven by Big Data in C/S Architecture, Computing, Performance and Communication Systems, 2024, Vol 8(1), 92-98

[13] Wang, Yuxin "Research on Intelligent Macro Image Recognition Algorithm of Oil Pipe Failure Based on Deep Learning." Journal of Image Processing Theory and Applications (2025), 8(1): 1-7

[14] Peng, Yicheng. "Construction and Evaluation of Credit Risk Early Warning Indicator System of Internet Financial Enterprises Based On AI and Knowledge Graph Theory." Procedia Computer Science 243 (2024): 918-927.

[15] Fan, Yijiao "Research on Risk Spillover Measurement of Fintech System Based on DCC-GARCH and Generalized Variance Decomposition Network Model."Accounting, Financial Engineering and Risk Management (2024), 7(5): 43-50

[16] Zhang J. Research on Fault Prediction and Health Management System of Railway Tunnel Drilling and Blasting Construction Machinery Based on Machine Learning[J]. International Journal of New Developments in Engineering and Society, 2024, 8(5),70-75

[17] Liu, Boyang. "Data Analysis and Model Construction for Crew Fatigue Monitoring Based on Machine Learning Algorithms." optimization 8.5: 48-52.

[18] Zhang, Yiru. "Design and Implementation of a Computer Network Log Analysis System Based on Big Data Analytics." Advances in Computer, Signals and Systems,(2024) 8(6),40-46.

[19] Xu, Hanyue. "Research and Practical Analysis on the Improvement Methods of Supply Chain Finance for Small and Medium-Sized Enterprises Based on Blockchain Technology." Financial Engineering and Risk Management 7.3 (2024): 118-123.

[20] Fan, Yijiao "Optimization Strategy and Implementation Effect Analysis of Unstructured Data Audit for Internal Audit of Commercial Banks in the Big Data Environment."Accounting, Auditing and Finance (2024), 5(3): 7-12

[21] Yu, Liu "Quantitative Analysis method in Accounting estimation Audit based on risk Identification Model and its Application in Audit quality improvement."Creative Economy (2024), 8(2): 43-48
[22] Zhang J. Research on Dynamic Stability Identification and Early Warning System for Engineering Vehicles Integrating Machine Learning and Data Driven Technology[J]. Academic Journal of Computing & Information Science, 2025, 8(2): 51-55.
[23] Zhu P. Construction and Experimental Verification[C]//Cyber Security Intelligence and Analytics: The 6th International Conference on Cyber Security Intelligence and Analytics (CSIA 2024), Volume 1. Springer Nature, 2025, 1351: 391.
[24] Tang, Zihan. "Research on Resource Planning and Transmission Line Optimization System of US Power Grid Based on GIS Technology." safety 8.5 (2024): 53-59.
[25] Huang, Tianyou. "Construction and Application of Financial Risk Warning System for Third-Party Payment Enterprises Based on Z-Score Model and Efficacy Coefficient Method." Academic Journal of Business & Management 6.12 (2024): 121-126.
[26] Wang Y. Research on Intelligent Diagnosis Technology of Oil Well Condition Based on Big Data and Deep Learning Model OWDNet[J]. Academic Journal of Architecture and Geotechnical Engineering, 2025, 7(1): 17-22.
[27] Dong, Peng. "Research and Analysis of Financial Crisis Prediction Model Based on the Fusion of Financial and Non-Financial Data with CSO Algorithm." Academic Journal of Business & Management 6.11: 56-60.
[28] Zhao, Yunpeng "Research on Computer Aided Information Analysis Technology Based on Data Mining and Social Network Analysis."Information and Knowledge Management (2025), 6(1): 8-13
*[29] Yang, **zhu. "Research on the Application of Medical Text Matching Technology Combined with Twin Network and Knowledge Distillation in Online Consultation." Frontiers in Medical Science Research, 2024, 6(11), 25-29*
[30] Peng, Yicheng. "Research on Exploring the Path of Financial Risk Management and Standardization Based on Artificial Intelligence Technology and Intelligent Investment Advisory Regulatory Framework." Financial Engineering and Risk Management 7.3 (2024): 110-117.
[31] Zihan Tang, Research on the Application of Joint Optimization Strategy of Energy Storage System and Demand Side Response Based on Tianqun Algorithm in Renewable Energy Grid, Frontiers in Power and Energy Systems, 2024, 3(1), 65-71
[32] Wang, Yuxin. "Application and Practice of Sensor Network Based on Deep Learning in Condition Monitoring of Underground Oil Production Equipment." International Journal of Frontiers in Engineering Technology 6.6 (2024).
[33] Jinshuo Zhang, Research on Intelligent Power Electronic Inverter Control System Based on Knowledge Base and Data Driven, Journal of Electrotechnology, Electrical Engineering and Management (2024), 7(3), 69-74
[34] Zhao F. Application and Performance Improvement of K-Means Algorithm in Collaborative[C]//2025 International Conference on Intelligent Systems and Computational Networks (ICISCN). IEEE, 2025: 1-6.
[35] Wang Y. Quality Control and Preventive Maintenance Site Management of Oil Drilling Machinery Equipment Based on Intelligent Monitoring[J]. Academic Journal of Engineering and Technology Science, 2025, 8(1): 50-55.
[36] Fan Y. Credit Rating Optimization Model Based on Deep Q-Network[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 464-475.

[37] Liu, Yu. "Build an Audit Framework for Data Privacy Protection in Cloud Environment." Procedia Computer Science 247 (2024): 166-175.
[38] Xu, Hanyue. "Analysis and Prediction of Financial Stock Risk Value Based On Improved FAST-ICA Algorithm and GARCH Model." Procedia Computer Science 243 (2024): 490-495.
[39] Zhang, Jinshuo "Optimization of Highway Engineering Design and Data-Driven Decision Support Based on Machine Learning Algorithm." Journal of Network Computing and Applications (2025), 10(1): 8-13
[40] Dong, Peng. "Research on Corporate Bankruptcy Prediction Combining Financial Data and Algorithmic Models Based on the Impact of Deleveraging." Accounting and Corporate Management, 2024.6(4),76-81
[41] Peng, Yicheng. "Development and Design of an Intelligent Financial Asset Management System Based on Big Data Analysis and Kubernetes." Procedia Computer Science 243 (2024): 482-489.
[42] Wang Y. Design and Implementation of a General Data Collection System Architecture Based on Relational Database Technology[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 561-572.
[43] Boyang Liu, Research on Holographic Retrieval and Analysis System for Scientific Research Data Based on SSH Framework and Lucene Engine, Journal of Artificial Intelligence Practice, 2024, 7(3), 164-171.
[44] Yang J. Application of Multi-model Fusion Deep NLP System in Classification of Brain Tumor Follow-Up Image Reports[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 380-390.
[45] Dong P. Construction of Movie Knowledge Graph and Design of Recommendation System Based on Movielens Dataset Expansion[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 540-550.
[46] Fan, Yijiao. "Research on Data Asset Feature Analysis and Value Evaluation Strategy Based on Random Forest and BP Neural Network." Academic Journal of Business & Management 6.11 (2024): 42-46.
[47] Zhang Y. Research on Optimization of Engineering Cost Database Based on Big Data and Intelligent Technology[J]. International Journal of New Developments in Engineering and Society, 2024, 8(5).42-47
[48] Liu, Boyang. "Design and Application of Experimental Data Management System Integrating Remote Monitoring and Historical Data Analysis." Journal of Electronics and Information Science 9.3 (2024): 160-167.
[49] Zhao, Fengyi. "Risk Assessment Model and Empirical Study of in Vitro Diagnostic Reagent Project Based on Analytic Hierarchy Process." International Journal of New Developments in Engineering and Society 8.5 (2024), 76 82 (2024).
[50] Zhang, Jinshuo "Research on Real Time Condition Monitoring and Fault Warning System for Construction Machinery under Multi Source Heterogeneous Data Fusion." Journal of Engineering Mechanics and Machinery (2024), 9(2): 139-144
[51] Huang T. Design and Implementation of a High Concurrency Online Payment Platform Based on Distributed Microservice Architecture[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 551-560.