Intelligent Analysis and Visualization of Loan Approval Based on Spring Boot

Yifan Guo[1], Jingjing Huo[1], Qingfeng Zhou[2]
[1]*College of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China*
[2]*iFLYTEK Co., Ltd., Hefei, Anhui, China*

Abstract: Aiming at the pain points of low efficiency, subjective decision - making and lagging risk control in traditional loan approval systems, this study designs and implements an intelligent analysis system for loan approval. The system uses technologies such as Spring Boot and ECharts to build a multi - dimensional data visualization platform covering visual analysis of loan approval, intelligent decision-making, user profiling, risk assessment, etc. Tests show that the system can significantly improve approval efficiency, reduce subjective decision -making bias, strengthen full - process risk management and control, and provide personalized services for customers. This system provides a low-cost, highly adaptable and easily implementable digital solution for small and medium - sized financial institutions, and has positive practical significance for promoting the transformation of the financial industry from "experience-driven" to "data - driven".

Keywords: Loan Approval; Data Visualization; Intelligent Decision - Making; Risk Assessment; User Profiling

1. Introduction

Under the guidance of the "Financial Power" strategy, fintech has become the core engine for promoting the formation of new productive forces in the industry [1]. Currently, the scale of China's consumer finance market is growing steadily, and the proportion of digital loan approval has exceeded 75%. However, the traditional approval model still has problems such as low efficiency and lagging risk early - warning, which are difficult to meet the diverse customer needs and regulatory requirements. The practice of commercial banks' bond investment research and trading business has confirmed that digital finance technology can significantly improve the accuracy and operational efficiency of financial business [2].
This study builds a digital platform covering the entire loan process based on big data and machine learning technologies, integrating functions such as visual display, intelligent analysis, and risk assessment, to provide small and medium - sized financial institutions with an efficient, accurate, and low - cost approval solution, and help the industry transform from "experience - driven" to "data - driven" [3].

2. System Design Requirement Analysis

(1) Business efficiency requirement analysis: The traditional manual loan approval cycle is 3-7 days, which cannot meet the financing timeliness requirements of small and micro - enterprises, nor is it suitable for the business scale of the consumer finance market exceeding 40 trillion yuan. There is an urgent need for the system to achieve automated approval to shorten the cycle and improve efficiency.

(2) Refined risk management requirement analysis

Refined risk management requirement analysis: Traditional approval relies on experience, resulting in a high difference rate in the approval of homogeneous customers and potential compliance risks. The difficulty in integrating multi - source data leads to one - sided pre - loan evaluation and lagging in - loan monitoring. The system needs to meet the requirements of regulatory compliance and risk traceability to achieve multi - dimensional real - time risk control.

(3) Customer personalized service requirement analysis

Customer personalized service requirement analysis: Financial institutions need to build user profiles based on customers' credit, loan purposes, etc., to achieve accurate customer segmentation and classification, provide personalized product recommendations and service plans, and promote the implementation of inclusive finance.

3. Introduction to Loan Approval Data

3.1 Data Introduction

This study integrates multi - source basic data from small loan websites and financial institutions [4] to support data analysis and model construction related to loan approval.

The data set contains 45,000 valid records and 14 core variables, comprehensively covering four key areas: demographic characteristics, economic status, loan information, and risk indicators. The variable design takes into account both comprehensiveness and pertinence. Demographic characteristics include the applicant's age, gender, and highest education level; economic status includes annual income, years of work experience, and property status; loan information involves the application amount, purpose, interest rate, and loan - to - income ratio; risk indicators [5] include the length of credit history, credit score, and historical default records. The approval result (approved / rejected) is used as the binary classification target variable.

3.2 Data Processing

Data collection uses a combination of batch upload and distributed storage. After being accessed through a Linux server, the data is synchronized to the HDFS distributed file system. The storage layer uses a hybrid architecture of "distributed storage + relational storage". HDFS stores the original and cleaned data, and MariaDB stores the structured analysis results. Data processing uses Spark as the core engine, and data cleaning and transformation are implemented through the Scala language. The DataX tool is used to synchronize data between HDFS and MariaDB. Finally, multi - dimensional aggregate statistics are realized through Spark SQL, and the MLlib is used to support the construction of the risk assessment model, providing a data basis for loan approval decisions.

4. Visual Analysis

The system uses ECharts to develop multi - dimensional visual chart functions, provides real - time data through the Spring Boot backend interface, and uses multiple charts to display the rules of loan approval, providing data support for business decision - making and risk control.

4.1 Visualization of Approval Distribution by Gender, Age, and Education

In traditional loan approvals, it is difficult for manual review to accurately identify the differences among customer groups in the cross - dimensions of age, gender, and education, leading to the loss of high - quality customer groups or misjudgment of high - risk customer groups. Figure 1 visually presents the application and approval rules of multi - dimensional populations.

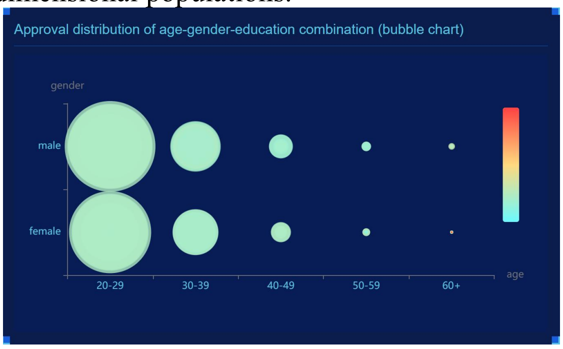

Figure 1. Bubble Chart of Combined Approval Distribution

As shown in Figure 1, the male undergraduate group aged 20 - 29 has the largest bubble, which is the core applicant group. The bubbles of high - education groups are concentrated in the age range of 25 - 35, and the approval rate of low - education groups is more stable after the age of 40. It is recommended that financial institutions target the male undergraduate group aged 20 - 29 as the key customer acquisition object, and strengthen pre - loan credit verification for young low - education groups to optimize resource allocation, improve approval efficiency, and enhance the accuracy of risk control.

4.2 Visualization of Risk Associated with Income and Loan Ratio

In traditional risk management, the correlation analysis between income and loan ratio is vague. Figure 2 intuitively presents the loan risk distribution of different income groups.

In Figure 2, the redder the color, the higher the risk. It can be seen that the risk of the group with an income of less than 20,000 yuan increases sharply when the loan ratio exceeds 20%. The risk of the group with an income of 20,000 - 50,000 yuan is relatively high when the loan ratio is 40% - 60%. It is recommended to strengthen risk - control review and limit loan amounts for the two high - risk customer groups: those with an income of less than 20,000 yuan and a loan ratio exceeding 20%, and those with an income of 20,000 - 50,000 yuan and a loan ratio of over 40%, to achieve precise risk management.

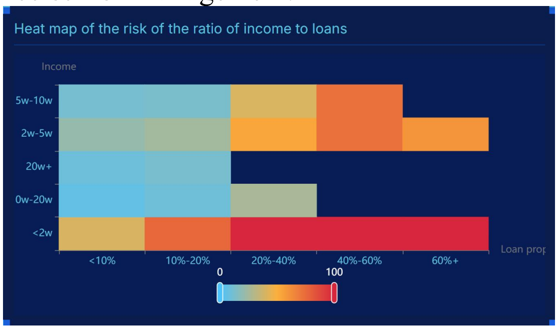

Figure 2. Heat Map of Risk Associated with Income and Loan Ratio

4.3 Visualization of Analysis of Loan Purposes

In the research on the differentiation of credit products, "the heterogeneous impact of loan purposes on application scale, approval volume, and amount" is an important practical issue. Figure 3 presents the differences in applications and amounts of different loan purposes from three dimensions: indicator type, loan - purpose type, and numerical value.

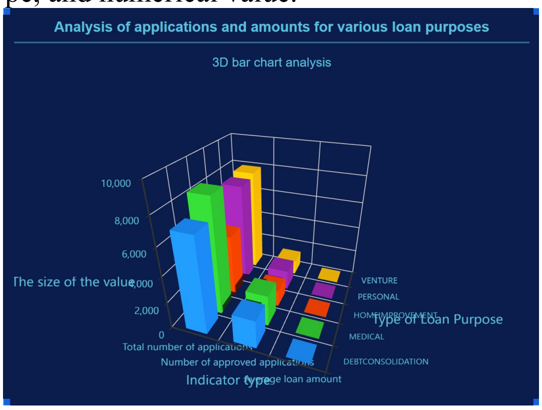

Figure 3. 3D Column Chart of Analysis of Loan Purposes in Terms of Applications and Amounts

From the total number of applications and approved applications in Figure 3, there is a significant positive correlation between the number of applications and the number of approvals, reflecting the transmission characteristics in the scale dimension. The VENTURE indicator shows the value advantage of single - business amounts with a high average loan amount. This bias is driven by the attribute differences of loan purposes (for example, business loans usually require large amounts of funds, while personal consumption/medical loans have more scattered needs), intuitively presenting the strategic bias characteristics of different loan purposes in the "scale - amount" dimension.

4.4 Visualization of Loan Application Conversion

In the research on the conversion efficiency of credit business, "the common customer - loss paths in different age groups during each conversion stage" is a core practical issue. Figure 4 presents the hierarchical conversion structure from consultation, pre - approval, formal application, to final approval, intuitively showing the "funnel effect" [4] through the decreasing absolute numbers, clearly depicting the customer - loss nodes from the initial consultation to the final approval, and providing a quantitative basis for analyzing the key bottlenecks in the overall conversion efficiency of each age group.

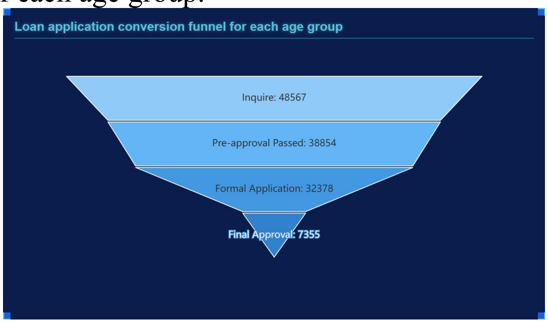

Figure 4. Funnel Chart of Loan Application Conversion in Different Age Groups

4.5 Visualization of Multi - Indicator Analysis by Age Group

In the research on the segmentation of credit customer groups, "the systematic impact of age segmentation on multiple loan indicators" is a key practical issue.

Figure 5 intuitively presents the distribution differences of each indicator among different age groups through color gradients. The number of loan applications in the age group of 20 - 29 is 32,378, which is the highest among all age groups and is the absolute main force in loan applications. The number of applications decreases step - by - step with age, while the average loan amount increases step - by - step.

The average loan amount of those over 60 years old is $12,100, ranking first. The approval rate and average credit score are significantly positively correlated with age. The interest rate of those over 60 years old is 11.3%, slightly higher, and the approval rate is the highest, indicating that age and credit score have a comprehensive impact on loan approval and interest rate pricing.

Figure 5. Heat Map of Multi - indicators by Age Group

Figure 6 conducts multi - dimensional synchronous comparisons of the characteristics of age groups such as 20 - 29, 30 - 39, and 40 - 49 from the dimensions of the number of applications, loan amount, average interest rate, approval rate, and average value. Through the shape of the radar polygon, it intuitively shows the relative advantages and disadvantages of each age group in different dimensions. For example, the 20 - 29 - year - old group stands out in terms of the number of applications, while other groups perform differently in dimensions such as interest rate and approval rate.

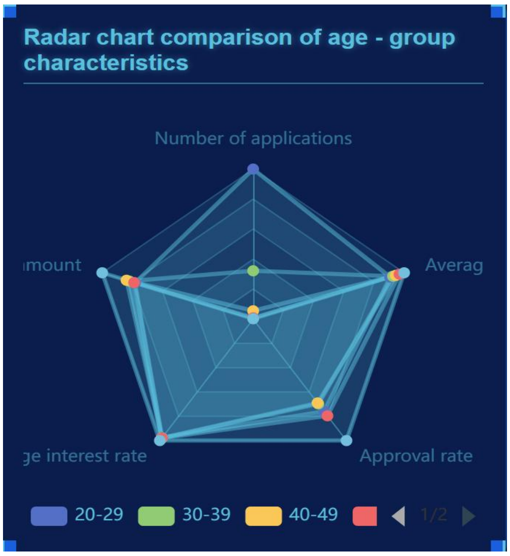

Figure 6. Radar Chart of Characteristics of Different Age Groups

5. System Design and Implementation

5.1 Intelligent Decision - Making

The intelligent decision - making module, as the core of the intelligent analysis system for loan approval, is based on the concept of data - driven, integrating artificial intelligence and big data technologies. With the Stacking integrated learning framework as the core, adopting a hierarchical architecture and microservice design [6], it constructs a closed - loop system of "data entry - intelligent analysis - decision output - full - process control" to achieve quantitative, automated, and traceable approval.

(1) Decision Panel: It supports the entry of applicant information and the upload of credit reports. After the front - end form verification ensures data compliance, the back - end algorithm engine is started. The random forest algorithm integrates and votes through multiple decision trees, learns features such as credit history and consumption behavior, captures non - linear associations, and completes the preliminary assessment of credit risk. The XGBoost algorithm iteratively trains to fit the relationship between income level, job stability, and repayment ability, and realizes the quantification of repayment ability.

After weighted integration by the logistic regression meta – learner [7], three types of decision suggestions and loan amounts are output, and the interpretability is enhanced in a visual way. The implementation effect is shown in Figure 7.

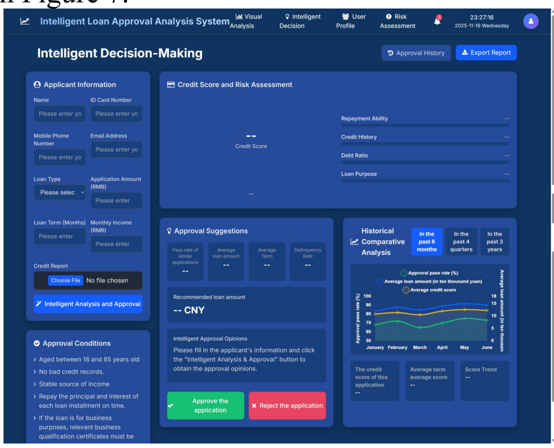

Figure 7. Intelligent Decision - Making Panel

(2) Application management: As shown in Figure 8, it realizes full - process automated control and supports batch processing and status tracking. Through standardized interfaces, it links each module [7] to track the lifecycle status of applications from submission to

disbursement, solving the problems of poor information transmission and redundant manual intervention in the traditional process and improving control efficiency.

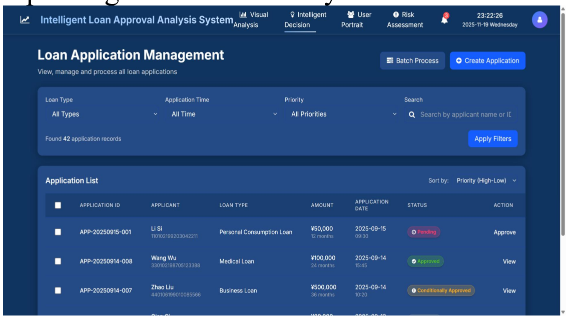

Figure 8. Application Management

(3) System Settings: It allows for the dynamic adjustment of parameters such as algorithm weights and risk - control thresholds to adapt to diverse business and regulatory requirements, enhancing the system's flexibility and applicability, as shown in Figure 9.

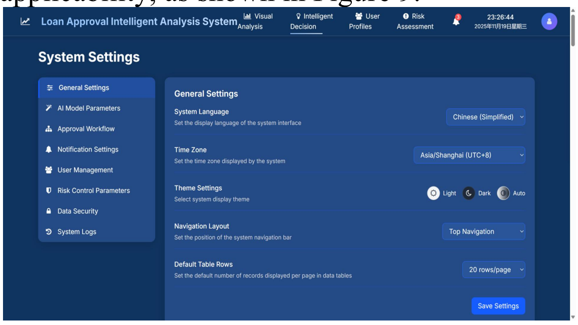

Figure 9. System Settings

This module, through the in - depth integration of functions and algorithms, helps financial institutions strengthen risk control, reduce costs, and increase efficiency, and becomes a key engine for digital transformation.

5.2 User Profiling

The integration of big data and artificial intelligence in financial risk control requires the establishment of a "hierarchical risk control + dynamic monitoring" mechanism. This system implements differentiated risk control in combination with user profiling [8]. User profiling is the key to achieving precise operation. Through data clustering, four typical types of users can be identified, and their characteristics and strategies are shown in Table 1.

(1) High - credit and Stable Type (28%): With a credit score of over 700, an income of 10,000 - 20,000 yuan, and a loan ratio of less than 30%, the risk is extremely low. A strategy of increasing the loan limit can be adopted to strengthen their stickiness as the core revenue - generating customer group.

(2) Young and High - potential Type (23%): Aged 25 - 35 years old, with a credit score of over 600, an income of 5,000 - 15,000 yuan, and active in applications. To meet their requirements for efficiency and flexibility, a media promotion strategy can be adopted, combined with loans tailored to professional scenarios, and accurately reach them through social channels to improve conversion efficiency.

(3) Dormant High - income Type (12%): With an income of over 25,000 yuan, a loan ratio of less than 10%, and a medium - high credit score but low activity. An interest - rate discount strategy can be used to target and wake them up, exploring the potential for value growth.

(4) Marginal Convertible Type (9%): With a credit score of 580 - 600, a loan ratio of 40% - 50%, and a qualified income. Since they fail to pass the traditional risk - control assessment, a risk - control optimization strategy can be adopted, introducing alternative data to improve the model and expand the customer base.

In summary, the hierarchical operation system for these four types of users helps the platform accurately allocate resources and maximize user value.

Table 1. Classification Table of Typical User Profiles

User Type	Proportion	Suitable Strategy
High - credit and Stable Type	28%	Increase Loan Limit
Young and High - potential Type	23%	Media Promotion
Dormant High - income Type	12%	Interest - rate Discount
Marginal Convertible Type	9%	Risk - control Optimization

5.3 Risk Assessment

Combining the platform's business characteristics and industry practice, the core risks are concentrated in three dimensions: credit risk, information security risk, and systematic risk [9]. The specific assessment is as follows:

(1) Credit Risk: It is the primary risk, accounting for 55%. There are significant risk differences among different user groups. Marginal convertible users have a high loan ratio and a relatively low credit score, with a relatively high probability of default. Some

users have the problem of insufficient information disclosure, which may easily lead to the platform's misjudgment of repayment ability and further increase the bad - debt rate.

(2) Information security risk accounts for 30%. The platform stores a large amount of sensitive user data such as identity and income, making it an easy target for cyber – attacks [10]. Technical vulnerabilities or insufficient internal control may lead to data leakage, which not only damages user rights and interests but also reduces the platform's trustworthiness and affects the continuous development of the business.

(3) Systematic risk accounts for 15% and has strong conductivity. In an economic downturn, the default rate of users is likely to increase. Coupled with industry fluctuations, it may cause liquidity pressure. If there is a risk event in an associated platform, it may trigger a panic among users to withdraw funds, which will impact the operational stability of the platform.

These three types of risks are inter - related and mutually amplifying. A dynamic monitoring mechanism needs to be established, and hierarchical risk control should be implemented in combination with user profiles to provide a basis for formulating subsequent risk - prevention strategies.

6. Conclusion

This paper focuses on the industry pain points of low loan approval efficiency and insufficient risk control accuracy, and designs and implements an intelligent analysis system for loan approval that integrates big data processing and visualization technologies, aiming to solve the problems of strong manual dependence and one - sided risk judgment in traditional approval. Practice shows that the system significantly shortens the approval cycle, with the proportion of automated processing exceeding 80%, significantly reduces the misjudgment rate of high - risk customers, achieves precise customer segmentation and matching of personalized services, and provides a feasible solution for financial institutions to reduce costs, increase efficiency, and comply with risk control. In the future, multi - modal large models can be integrated to expand scenarios such as intelligent due diligence and real - time anti - fraud, and deepen the service ability of inclusive finance.

References

[1] Zhu Keli. "Financial Powerhouse: One of the Five Major Articles": Driven by Science and Technology Finance, Accelerate the Formation of New Qualitative Productivity. Financial View, 2024(1).

[2] Research Group of the Fund Operation Center of Agricultural Bank of China, Shi Wei. Practice and Reflection on Digital Finance Assisting Commercial Banks to Improve Bond Investment Research and Trading Abilities. Bonds, 2025, (10): 76 - 81.

[3] Wang Xinru, Ding Jiashu, Zhao Wei, etc. Data - Driven Decision - Making: Intelligent Talent Evaluation Center Empowers Human Resources Management. Enterprise Management, 2025(11): 1 - 2.

[4] Wu Xuchuan. Application and Role of Big Data in the Financial Field. Heilongjiang Finance, 2018, (08): 16 - 18.

[5] Zhang Jiaqian, Li Wei, Ruan Sumei. Prediction of Loan Default Risk Based on Machine Learning. Journal of Changchun University of Science and Technology (Social Science Edition), 2021, 34(4): 105 - 111.

[6] Zhao Zhihong, Du Ying. Research on Credit Fraud Risk Prediction Model Based on Machine Learning. Modern Electronics Technique, 2021, 44(24): 177 - 181.

[7] Shen Yawen. Research on Credit Risk Evaluation of Technology - Based Small and Medium - Sized Enterprises in Bank A. South China University of Technology, 2019(01): 27 - 38.

[8] Liu Zheng, Zhao Gang, Zhang Xiang. Analysis of the Integration and Practice of Big Data and Artificial Intelligence in Financial Risk Control//Proceedings of the 2025 Academic Exchange Conference on Social Development and Innovation. 2025: 1 - 3.

[9] Chen An, Chen Ning, Fan Chao. A Review of Data Mining Techniques in Financial Credit Risk Evaluation. Intelligent Computer and Applications, 2017, 7(5): 55 - 59.

[10] Li Yichen, Zheng Zijia, Miao Jiafang, etc. Application of Artificial Intelligence in Financial Risk Prevention and Control in the Big Data Era. Fortune Today, 2023(11), 98 - 97.