Wang, Xiaoning; Tang, Yi; Quaranta, Anna Grazia

Periodical Part

Machine learning-driven lending decisions in bank consumer finance

International journal of information systems and supply chain management

Provided in Cooperation with:
ZBW OAS

Reference: In: International journal of information systems and supply chain management Machine learning-driven lending decisions in bank consumer finance 17 (2024).
https://www.igi-global.com/ViewTitle.aspx?TitleId=348337&isxn=9798369324738.
doi:10.4018/IJISSCM.348337.

International Journal of Information Systems and Supply Chain Management
Volume 17 • Issue 1 • January-December 2024

Machine Learning-Driven Lending Decisions in Bank Consumer Finance

Xiaoning Wang
Chongqing Institute of Engineering, China

Yi Tang
Chongqing Institute of Engineering, China

Anna Grazia Quaranta
University of Macerata, Italy

ABSTRACT

This paper investigates the bank lending decision process for internet consumer finance using machine learning. It focuses on microloans and compares Logistic Regression and GBDT models for credit risk assessment. Variables are filtered and recoded via Information Value and WoE methods to enhance discrimination between defaulting and performing users. Experimental results utilizing these models predict credit risk and optimize using AUC values. Additionally, it develops a fixed-effect regression model to explore how bank-specific factors affect systemic risk, revealing that larger banks reduce risk, while higher returns, non-performing loans, and equity volatility elevate it, with inconclusive effects from leverage ratio.

KEYWORDS

Machine Learning Algorithms, Bank Financing, Lending Decision Mechanism

In recent years, with the continuous development of the Internet, its impact on traditional industries has been almost ubiquitous, and the combination of the Internet and its related technologies with the financial sector is gradually affecting the traditional financial industry and gradually having a transformative impact on the fields of money payments and money financing. The knotting of Internet technology with traditional consumer finance has then led to the creation of Internet consumer finance (Sima et al., 2020). In simple terms, Internet consumer finance is a credit activity in which small loans are provided by Internet finance companies to meet individual consumers' consumption needs for goods and services and are repaid in installments (Xiao & Tao, 2021). Compared with traditional consumer finance, Internet consumer finance has the characteristics of convenient and fast services, substantially lower transaction costs, and wider coverage groups. With the continuous development of the economic level and the upgrading of people's consumption concepts, Internet consumer finance is also gradually being recognized by more consumers (Pahlevan & Naghavi, 2020).

Today's Internet consumer finance market mainly has the following three product forms. First is relying on its Internet financial platform for consumers to provide small consumer loans and installment shopping services. Secondly, this study examines the installment shopping platform within the Internet consumer financial services model. This platform allows consumers to apply for installment purchases or small consumer loans. It provides product brands for installment shopping users and facilitates the application process for consumers.(Chen et al., 2021). Third is the Internet consumer financial services model of banks, i.e., traditional banks' personal consumer loan business

DOI: 10.4018/IJISSCM.348337

(Wu & Zhang, 2020). This study used machine learning to model and analyze user data from Internet credit platforms, offering targeted strategies for credit risk assessment within the constraints of limited computing resources and real-time responsiveness, given the necessity for comprehensive user profiling in the current Internet finance landscape. In addition, this paper explores the specific implementation of credit risk prediction techniques, which is also informative in this regard.

The development in the field of consumer finance presents obvious differences between the Internet and traditional models. Internet consumer finance takes online platforms as the main operation channel, using Internet technology to realize efficient and convenient financial services, with diversified products, simplified approval process, and emphasis on user experience and convenience; while traditional consumer finance relies on entity institutions, with products including credit cards and personal loans, which have advantages in credibility and trustworthiness, though the approval process is more cumbersome.

In terms of business model, Internet consumer finance adopts technical means such as online application, automatic approval, and fast lending to reduce operating costs, emphasizes connecting borrowers and lenders, and provides information services and risk management. In contrast, traditional consumer finance relies on entity institutions to provide services, and the approval process is relatively cumbersome but has a perfect internal control and risk management system. In terms of risk control and supervision, Internet consumer finance faces challenges such as information security, privacy protection, and capital risk, and is subject to strict supervision; while traditional consumer financial institutions are relatively mature in risk control and supervision and are subject to strict supervision by the traditional financial regulatory system. In terms of market competition and innovation, the Internet consumer finance market is highly competitive and innovative, continuously improving its competitiveness through technological innovation and service innovation to meet consumers' diversified financial needs; while traditional consumer financial institutions are actively transforming in the face of the competitive pressure of Internet consumer finance, strengthening digitalization and innovating their services to enhance their market competitiveness.

The economy is the basis of finance, and the financial industry is the core of the modern economy, which not only plays an important role in macro-control but also penetrates every aspect of human life. As the main component of the financial industry in China, in addition to the three major policy banks, the dominant role of commercial banks is very prominent, especially the large state-owned commercial banks (Ziemba et al., 2021). Most enterprises can raise relatively low-cost funds from commercial banks, and this mode of indirect financing will not undergo a fundamental change in a certain period of time (Panori et al., 2019). At the same time, for commercial banks, public credit assets are also their indispensable foundation, especially the credit business of large companies, which, to a considerable extent, affects the business development and profit realization of commercial banks.

As far as the current public credit risk management of commercial banks is concerned, it is far from that of commercial banks in developed Western countries in terms of management philosophy, organizational structure, and technical means. Because it is difficult for domestic commercial banks to make direct changes to the external macro-environment, the effective management of public credit risk depends on their efforts, and the improvement of this management level is also important for commercial banks to achieve their strategic objectives and shape their core competitiveness (Gharaibeh et al., 2017). Among them, *credit issuance review* is the review of credit businesses for lending. In the case of large companies, it refers to the process of the auditor giving credit instructions after reviewing the implementation of the credit conditions one by one based on the approved requirements for various credit projects that have passed the credit review and approval, including but not limited to the basic documents of the enterprise, the description of the financing resolution, the credit business contract, the actual credit use, the risk mitigation conditions, and other credit elements .

The consumer finance sector is at the intersection of rapid development and increased regulation, and its development faces many challenges and opportunities. First, data privacy protection has become one of the industry's concerns. With the extensive use of personal information in consumer finance activities, regulators have stepped up their supervision of data privacy protection, emphasizing that companies must strictly follow relevant regulations in the collection, storage, and use of personal information, or else they may face heavy penalties or even legal proceedings. Second, cybersecurity risks are escalating, threatening the safety of consumer data and funds. Malware, phishing, and other forms of cyberattacks continue to emerge, requiring financial institutions not only to strengthen their investment in cybersecurity and implement stricter security measures but also to enhance their ability to respond to cybersecurity risks. In addition, consumer protection is an important safeguard for the development of the industry.

By strengthening the supervision of financial product compliance and consumer rights and interests, regulators have prompted companies to operate in a lawful and compliant manner and enhanced the overall credibility of the industry. At the same time, industry self-regulation and regulatory cooperation are particularly important. The consumer finance industry should strengthen the construction of norms and self-regulation, and, at the same time, actively cooperate with regulators to jointly formulate industry standards and norms and build a good industry ecology. Finally, technological innovation and regulatory adaptation are the keys to the sustainable development of the industry. Financial institutions need to strengthen technical investment and continuously improve their technical level and innovation ability to adapt to changes in regulatory policies and market demand and maintain the competitive advantage of the industry. In summary, when facing regulatory challenges, the consumer finance sector needs to be supported by efforts in data privacy protection, cybersecurity risk prevention, consumer rights and interest protection, industry self-regulation and regulatory cooperation, and technological innovation and regulatory adaptability to achieve healthy and stable development of the industry.

In recent years, with the booming development of the Internet, the traditional financial industry has been greatly affected. In particular, the combination of Internet technology and finance is gradually changing the traditional way of monetary payment and financing. In this process, Internet consumer finance has attracted much attention, and its advantages of convenience and speed, low transaction costs, and wide coverage of groups are highlighted. At present, Internet consumer finance is mainly presented in three forms: first, relying on Internet financial platforms to provide small consumer loans and installment shopping services; second, the Internet consumer financial service model provided by installment shopping platforms; and third, the personal consumption loan business of traditional banks. Meanwhile, the use of machine learning technology to model and analyze the user data of Internet credit platforms in order to provide targeted credit risk assessment strategies has become an important research direction in the field of Internet finance.

However, commercial banks play an important role in the national economy. Large state-owned commercial banks, especially, have significant advantages in fund mobilization and credit business. To improve business development and profit realization, commercial banks need to strengthen public credit asset management. With regard to the credit risk control system of micro and small enterprises, the researcher puts forward some improvement measures and suggestions to strengthen risk management and promote the improvement and development of the overall credit risk management system of micro and small enterprises in the industry.

This paper proposes improvement measures for credit risk control systems tailored for micro and small enterprises within Industrial and Commercial Bank of China (ICBC) branches. Despite ICBC's robust framework for credit risk management, challenges persist, particularly for small and microcredits. Through analysis of ICBC's risk identification and prevention initiatives, this study offers insights to enhance overall credit risk management for micro and small enterprises. By examining ICBC's credit risk assessment for (micro- and small enterprises) MSEs, strategic insights are provided to expand financing options and alleviate funding difficulties, promoting MSE development and China's economic advancement. Additionally, financial engineering, including innovative financial technology applications, is explored to enhance financing channels for MSEs. Amid economic pressures, regulators have implemented strategies such as easing bank supervision and providing targeted assistance programs. Quantifying risk through macro-indicators and big data analysis is crucial for commercial banks to optimize lending practices while maintaining stability, especially amid pandemic factors.

LITERATURE REVIEW

Scholars have pointed out a distinction between Internet finance in a narrow sense and a broad sense, and the difference between the two lies in whether they include the Internet-based form of finance (Orlova, 2020). Bazarbash presented a similar view that Internet finance is a general term for financial activities in which financial service providers provide financial services in banking, insurance, and securities through an Internet platform, with Internet technology as the technical support (Bazarbash, 2019). *Internet finance*, in a narrow sense, refers to the establishment of a bridge between financial service providers and users through Internet technology, establishing a new financial operation model, while Internet finance, in a broad sense, includes not only the financial operation model but also the external environment such as Internet financial institutions and Internet regulation (Caird & Hallett, 2019). Specifically, some scholars' understanding of Internet finance mainly lies in the narrow scope of Internet finance, which is considered to be an innovation in financial forms. Among them, scholars believe that the "new finance" of the Internet is not a kind of financial exuberance but a change in the financial sales and access channels, whether online or offline; the transaction is a financial contract. Internet finance is probably just a stage in the development of the financial industry, a transitional form of the financial industry in the process of reform (Kakderi et al., 2019). Tsarchopoulos et al believe that Internet finance enhances the efficiency and user experience of financial services through Internet technology, broadens the new boundaries of financial services, and thus further promotes the reform of the financial sector but does not change the main functions of finance in the six areas of payment, price discovery, financing, risk management, resource allocation, and dealing with information asymmetry (Tsarchopoulos et al., 2017). In other words, Internet finance does not derive new financial functions but only improves the efficiency and experience of financial services.

The earliest risk models were based on modern investment theory. To give the model an optimal loan allocation decision, Monte Carlo simulation was used to obtain key parameters such as yield and standard deviation for each year of the loan term, artificially introducing various constraints, such as risk limits, adjusting the strategy, international practice of laws and regulations, and business management, and finally establish return maximization as the objective function and calculate it (Komninos et al., 2019). Starting from the financial crisis in 2008, the International Accounting Standards Board (IASB) established an expected credit loss model for expected loss (EL) under the Basel framework and proposed new impairment accruals such as International Financial Reporting Standards (IFRS) 9. The relevant research argues that the implementation of the new credit risk regulatory model, involving profit and loss and credit risk, requires the collection of data calculation systems distributed by two major departments of finance and risk management and operationally requires consideration of whether the forecasts of different asset-related parameters can meet the adequacy and reliability of historical data (Rana et al., 2019). Accordingly, scholars believe that the new regulation will affect the lending strategies of state-owned banks and joint-stock banks, suggesting that commercial banks should re-examine their market exposures, strengthen the dynamic early warning of DVO1 for "current profit and loss changes in financial assets at fair value," and encourage commercial banks to use interest rate derivatives for risk hedging (Trencher & Karvonen, 2020).

The banks should also encourage commercial banks to use interest rate derivatives to hedge their risks and require further strengthening of their credit rating capabilities for customers.

The key to preventing systemic risk lies in the correlation network among financial institutions, and researchers construct and test the degree of contribution of a single financial institution to systemic risk in the system by decomposing the spillover factor loadings through factor analysis measures and systemic risk factor attribution methods (Turkson et al, 2016). Several scholars have conducted indicator measurements using complex network theory to analyze the guarantee relationships among enterprises: from the perspective of guarantee, two indicators of steady-state risk density and risk propagation speed are constructed to evaluate the network's ability to resist risk, but with certain subjectivity of node selection and computational complexity (Zheng, 2022).

Social Impact of Internet Consumer Finance

In the development of financial technology (*fintech*), especially in the field of Internet consumer finance, in addition to the importance of technology and data analysis, ethical considerations and social impacts need to be taken into account. First, when banks use big data and algorithms for risk control, they need to ensure customer privacy and data security, comply with relevant laws and regulations, and take measures to safeguard the legitimate use and protection of customer data. At the same time, banks should establish transparent mechanisms for data use and algorithmic decision-making, explaining to customers how the models affect lending rates and churn, and ensuring that these decisions are fair and reasonable and do not discriminate against any particular group.

Second, banks should take into account social responsibility and sustainability considerations when conducting risk control operations. In particular, when granting loans to MSEs, in addition to focusing on the credit rating and default probability of the enterprise, they should also consider the social impact and sustainability of the enterprise. Banks can assess the social responsibility and sustainability of enterprises by establishing environmental, social, and governance (ESG) indicators, incorporating these indicators into loan approval and pricing considerations, and promoting enterprises to develop in the direction of sustainable development and make positive contributions to society.

In addition, when banks use technology for risk control, they should also be aware of the potential risks and negative impacts of technology. For example, algorithmic decision-making may be biased and discriminatory, leading to unfair treatment of specific groups; big data analysis may exacerbate information asymmetry, making it more difficult for certain enterprises or individuals to obtain loans; and intelligent risk control systems may exacerbate the volatility of the financial market and increase systemic risk. Therefore, banks need to establish a risk management mechanism to identify and respond to the risks brought by the technology in a timely manner and ensure that the development of fintech is in line with the overall interests of society and the economy.

In summary, the development of Internet consumer finance requires not only the support of technology and data analysis but also the consideration of ethical factors and social impacts to achieve a benign interaction between risk control and social responsibility and promote the sustainable development of the fintech industry.

Analysis of Lending Decision Mechanism of Bank Financing Category with a Machine Learning Algorithm

In the field of finance, it is crucial to construct effective risk control models. The aim of this paper is to discuss the model construction for default prediction of borrowing users and its performance evaluation and improvement process. For this binary classification problem, machine learning methods are used, of which logistic regression is a common algorithm. The model training process utilizes the observations of defaulting companies as positive samples, and the optimal features and their number are determined through iterative operations. To improve the model performance, the recursive feature elimination (RFE) algorithm was used to select the optimal set of features, and the model performance was evaluated using indicators such as receiver operating characteristic (ROC), area under the ROC curve (AUC), and sensitivity. In practical applications, the accuracy or error rate is balanced by setting probability thresholds to ensure accurate prediction of defaulting enterprises and avoid misjudging normal enterprises as defaulting. In addition, in bank loan decision-making, the special nature of microenterprise loan business and information asymmetry problems require the assistance of credit specialists in order to reduce the loan risk. Through the continuous adjustment and optimization of the model, the application effect of the model in the actual business is improved, to manage financial risks effectively.

In the financial services sector, the rise of Internet consumer finance has brought about many technological innovations to provide more convenient and personalized financial services. Big data and artificial intelligence technologies have enabled more accurate credit assessment of borrowers, blockchain technology has improved the transparency and security of transactions, and mobile internet technology has made financial services more convenient and accessible. However, the attendant risks and regulatory challenges cannot be ignored. Companies need to establish a sound risk management system, strengthen data security and privacy protection, and pay close attention to changes in regulatory policies, laws, and regulations.

In terms of user experience, Internet consumer finance focuses on personalized services and improves customer satisfaction and loyalty through intelligent recommendations and mobile application experience. At the same time, Internet consumer finance also promotes financial inclusion and financial inclusion, lowering the threshold of financial services and expanding the coverage of financial services, especially for those groups that are hard to reach by the traditional financial system. Internet consumer finance will continue to grow and develop. More innovative fintech products and services, such as virtual currency payments and smart contract lending, may emerge in the future, driving the financial industry towards digitization and intelligence. Meanwhile, the company may strengthen cooperation and integration with traditional financial institutions and technology companies to promote financial innovation and digital transformation jointly and even seek international expansion into overseas markets. However, the challenges of cross-border regulation and risk management will follow.

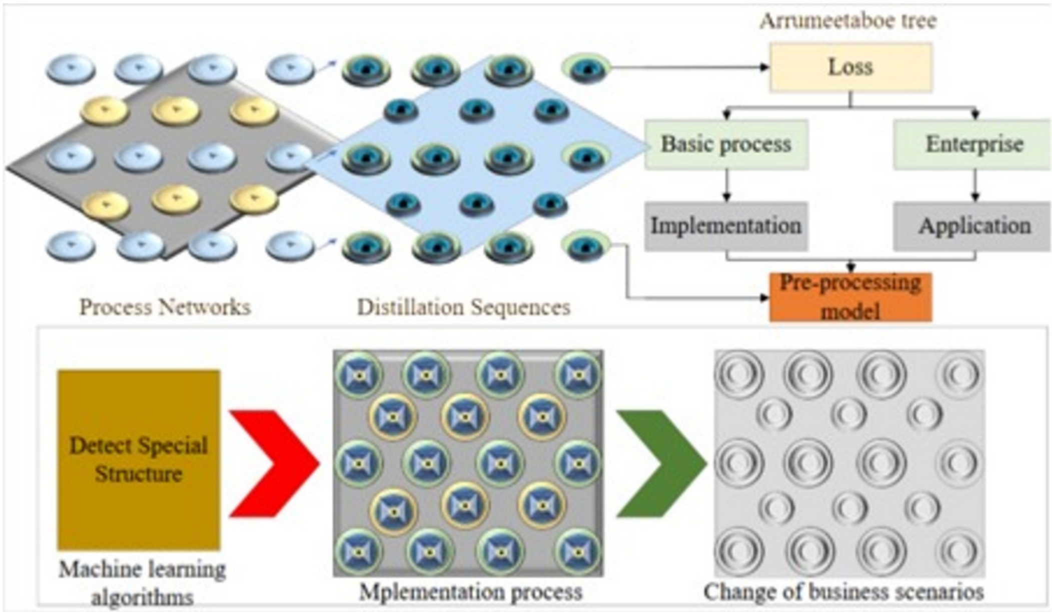
Improved Machine Learning Algorithms for Lending Decisions

Logistic regression is a common model for classification (Lee & Shin, 2020). In classification problems, one needs to predict whether a sample belongs to a certain category, for example, to determine whether an email is spam, to determine whether a financial transaction is financial fraud, and to determine whether a patient suffers from a certain disease based on the results of a medical examination. In the field of risk control, we often need to determine whether a loan user will default. Logistic regression models are widely used in the field of risk control in Internet finance because of their fast classification speed, suitability for binary classification problems, and interpretability (Mohanty et al., 2016).

Machine learning is a comprehensive discipline spanning several disciplines, such as computer science, engineering technology, and statistics, requiring expertise in multiple disciplines, and it can even be said that machine learning can be applied to any field where data needs to be interpreted and manipulated. In today's era, the emergence and development of electronic sensors and databases have made the data that humans can record and save increasingly huge, and even all aspects of our lives can be recorded by data of all shapes and sizes (Aphale & Shinde, 2020). Faced with such a huge amount of data, it would be very meaningful for all aspects of human life if regular patterns present in the data could be identified and understood through some algorithm or technique (Alzubaidi et al., 2021). The research area of machine learning is the invention of computer algorithms to transform data into intelligent behavior. The key to machine learning lies in identifying regular information or patterns in the data that are meaningful to human decision-making behavior through automated or semi-automated methods (Ziemba et al.,, 2021).

In elaborating on the concept of machine learning, both automated and semi-automated methods and information or patterns that are meaningful to human decision-making behavior must be emphasized. The former indicates that automated means must be used to process the data collected by various sensors and databases collected and stored in massive amounts, and the latter illustrates that machine learning must incorporate the understanding, exploration, and analysis of data rather than the application of an off-the-shelf algorithm or framework to accomplish a task. The implementation

Figure 1. Improved machine learning algorithm framework for lending decisions



of machine learning must, therefore, be based on a thorough understanding of the business or data requirements rather than simple and crude use of computer algorithms or techniques.

With the change of business scenarios, the implementation process of using machine learning algorithms to solve business problems may vary, but the basic process is the same in the enterprise implementation application, which mainly involves the process of data collection, pre-processing model development, and implementation. In this process, the most important thing in using machine learning to solve business problems is not the specific algorithm or the technology of data processing itself; the most important thing is whether the understanding of the business problem is thorough (Mehmood et al., 2017). If one does not proceed with a thorough understanding of the business and apply a model and algorithm directly to the data, the results will not be very good in many cases. Therefore, when using machine learning algorithms to solve business problems, it is necessary to communicate deeply with business departments to clarify whether machine learning methods need to be used to solve this business problem, which types of algorithms and models are more suitable for this kind of problem, and whether some business industry norms need to be followed.

All these issues must be fully communicated and solved before model construction is carried out, as shown in Figure 1. *Association analysis* is a technique used in the fields of data mining and machine learning to discover intrinsic connections hidden in massive amounts of data, which can be represented by association rules. The most classic example of the association rule algorithm is the "beer and diapers" case of Walmart, which found that most of the customers who bought diapers would also buy beer through association analysis, so the supermarket put beer and diapers together, thus increasing the sales of these two products (Angelidou, 2017). In addition to merchandising, correlation analysis is also used in many other areas, such as in the field of geosciences, where data on the atmosphere, hydrology, and geology of an area can be analyzed to reveal to geoscientists the intrinsic connections of the geoscientific nature of the area in question.

Two things need to be taken care of in correlation analysis: first, how to explore and reveal the connections between things, and second, how to make sure that the connections found in the previous step are not false connections so as not to produce false analysis results. See Equations 1–2.

$$X = \beta_1 x_1 - \beta_2 x_2 + \beta_3 x_3 - ,\ldots,\beta_m x_m \tag{1}$$

$$P(x) = p_i \tag{2}$$

This means that the liquidity of commercial banks has gradually shifted from a single asset side to an asset plus liability side, where commercial banks themselves can raise external financing and reuse the acquired funds to extend loans to enterprises, again earning a carry return. As a result, commercial banks have been able to maintain their asset size and adequate liquidity through non-owned funds, an important shift that is no less than a banking revolution. In the current economic environment, most microfinance companies are relatively loosely managed, have unsound financial data, and lack a sound disclosure system. At present, relatively few managed liability companies (MLCs) are successfully listed on the domestic Growth Enterprise Market (GEM) and National Stock Exchange (NSE), so many MLCs are turning to the offshore exchange market, where issuance conditions are relatively lenient. Although the overseas listing of members of parliament (MPs) is a good path, it also faces many difficulties: master limited partnerships (MLPs) need to make corresponding adjustments according to their shareholding structure before listing declaration and, at the same time, need to carry out a series of preparatory activities, such as setting up special purpose entities abroad, resulting in a relatively long preparation time for listing; the overseas stock exchange market has very strict requirements for information disclosure. The overseas stock exchange market has very strict information disclosure requirements, which requires MLCs listed overseas to comply strictly with the information disclosure system. MLCs listed overseas to raise funds also need to make a declaration to the State Administration of Foreign Exchange and other relevant departments and obtain approval before the funds can flow back to the territory, which also increases the difficulty of operation. See Equation 3.

$$H(Y|X) = \sum_{i=1}^{n} p_i H\left(Y \middle| X = x_i^2\right) \tag{3}$$

Based on this concept, we introduce the notion of information gain, which expresses the extent to which the information uncertainty of a class is reduced while the features are given. See Equation 4.

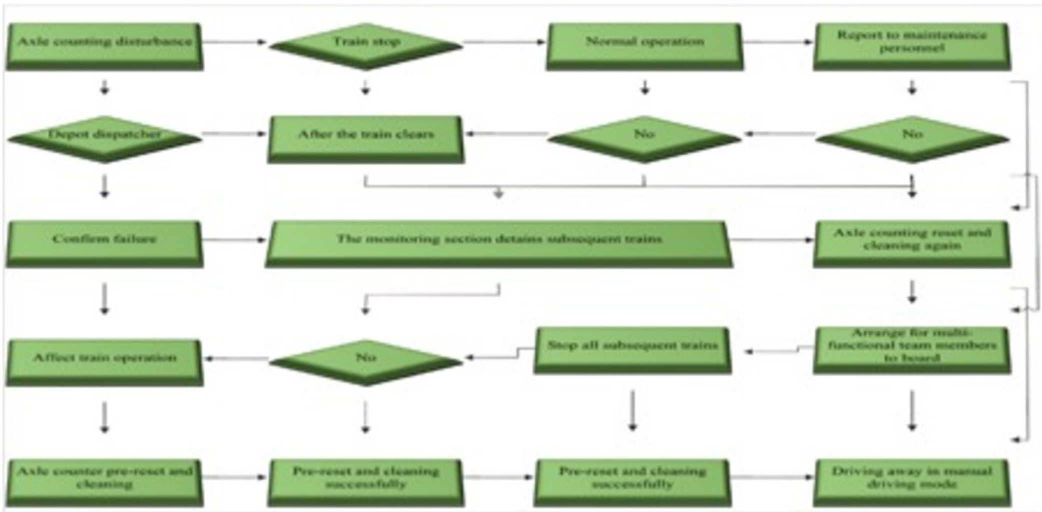
$$G(D,A) = H(D) + H(D|A) \tag{4}$$

In general, different features have different information gains, and the larger the information gain of a feature, the stronger its classification ability. Therefore, we usually choose the feature with the largest information gain for split construction. See Equation 5.

$$\sum_{i=1}^{n} p_i H\left(Y \middle| X = x_i^2\right) = D \tag{5}$$

The core of the algorithm is the recursive construction of the decision tree and the information gain is used as a criterion for feature selection at each node during the recursion until the recursion stop condition is satisfied. The final step is to prune the decision tree, which is mainly constructed by recursively branching each tree node until it can no longer be constructed. In this case, when the sample has more features, the generated decision tree tends to overfit the sample data, resulting in a model that scores very well on the training set but performs very poorly on the test set, i.e., the decision tree model tends to overfit and generalize poorly. The reason for overfitting is that the decision tree we build is too complex, causing it to overlearn the training samples and thus making the model

Figure 2. Bagging framework



have poor generalization ability. The solution to this problem is to simplify the generated decision tree and make it simpler. Decision pruning is the process of simplifying the generated decision tree. Specifically, we make the decision tree simpler and achieve better prediction on unknown test data by subtracting the leaf nodes that are too finely divided and making the parent node of the leaf node a new leaf node, thus improving the generalization ability of the model, as shown in Figure 2.

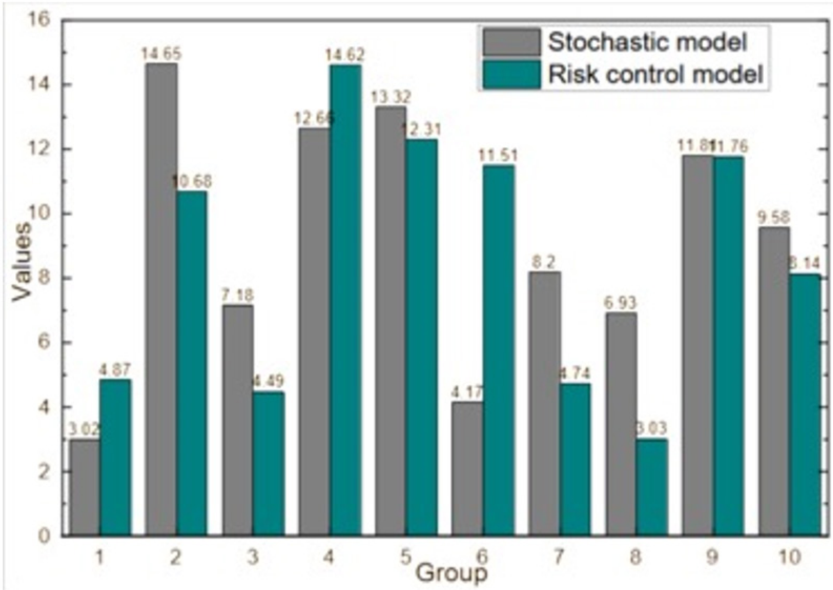
Understanding the decision tree algorithm and the idea and algorithmic logic of bagging, random forest is better understood. There are two main improvements that the random forest makes over bagging. The random forest algorithm uses a CART decision tree as a weak learner; based on the use of a weak learner, the random forest also makes improvements to the building of the decision tree. This discusses how random attributes are selected during the learning process of a decision tree, which serves as the base learner for constructing bagging integrated learning. For the ordinary decision tree model, we usually select an optimal feature among all n samples on the node as the left and right subtree division of the decision tree, while the random forest model selects some of the sample features on the node randomly. See Equations 6–7.

$$E(f,D) = \frac{1}{m}\sum_{i=1}^{m}\prod\left(f(x_i) = y_i\right) \tag{6}$$

$$A(f,D) = \frac{1}{m}\sum_{i=1}^{m}\prod\left(f(x_i) = y_i\right) - E(f,D) \tag{7}$$

The Internet financial risk control model we build needs to predict whether a borrowing user will default or not, which is a binary classification problem. For classification problems, although we commonly use error rate and accuracy, these two model performance metrics do not satisfy all task requirements. Take the problem of identifying defaulted users solved by the risk control model as an example. Suppose there is a group of users borrowing through the Internet financial platform. Obviously, the error rate measures what percentage of users are identified incorrectly, but if we care about what percentage of the picked users will not default or the percentage of all users who will

Figure 3. Model grouping performance



not default, then the error rate cannot be measured anymore, and we need to introduce other model performance metrics.

EXPERIMENT DESIGN

A bank's decision to lend begins with predicting the probability of a firm's default. Therefore, the machine learning model should use the observations of certain firms that have defaulted as a positive sample. The parameters need to train the classifier on the experimental dataset by initializing each feature and then iterative operations to find the optimal features and the number of feature inputs to the dataset. The features with a low contribution to the model prediction are pruned out in the current feature variables (Xie et al., 2019). The RFE greedy optimization algorithm is used as the objective function to find the set of features with the best model performance. The number of running result features can achieve the highest score at the 5th one. For the evaluation metrics of the dichotomous problem, the ROC and AUC evaluation metrics in machine learning are used, and the results of the modeling, where *sensitivity* (the proportion of positive values predicted by the model among all data with positive true outcomes) is greater than 0.95 for the first time. The logistic regression prediction accuracy of the model (predicting the category of all corporate behaviors: default or non-default status) reached 0.8, and the AUC reached 0.89, which represents a strong predictive power. In applying the model, setting a probability threshold of 0.408, based on information provided by historical data collected on firms' incoming and outgoing bills and ratings, accurately predicts 100% of the sample of firms in default but also incorrectly indicates that about 15% of normal firms are in default (model risk). The weights of the derivative variables in the logistic Stee regression model are considered in the business, as shown in Figure 3.

This problem occurs more frequently in the case of loans for fixed-asset projects. For example, some commercial banks granted loans to government purchase service projects that were not included in the budget without authorization. Certain road and railway, urban construction, and wind power generation projects used the Land and Resources Bureau's pre-approval of land instead of formal land approval or incomplete land approval or lacked project compliance elements such as construction

Table 1. Comparison of the results of the first training

Training method	*F*-metric misrecognition rate	AUC value	KS value
Logistic regression	13.25.	3.56.	8.51.
Neural networks	8.23.	9.5.	8.44.
Decision tree	4.96.	7.96.	2.62.

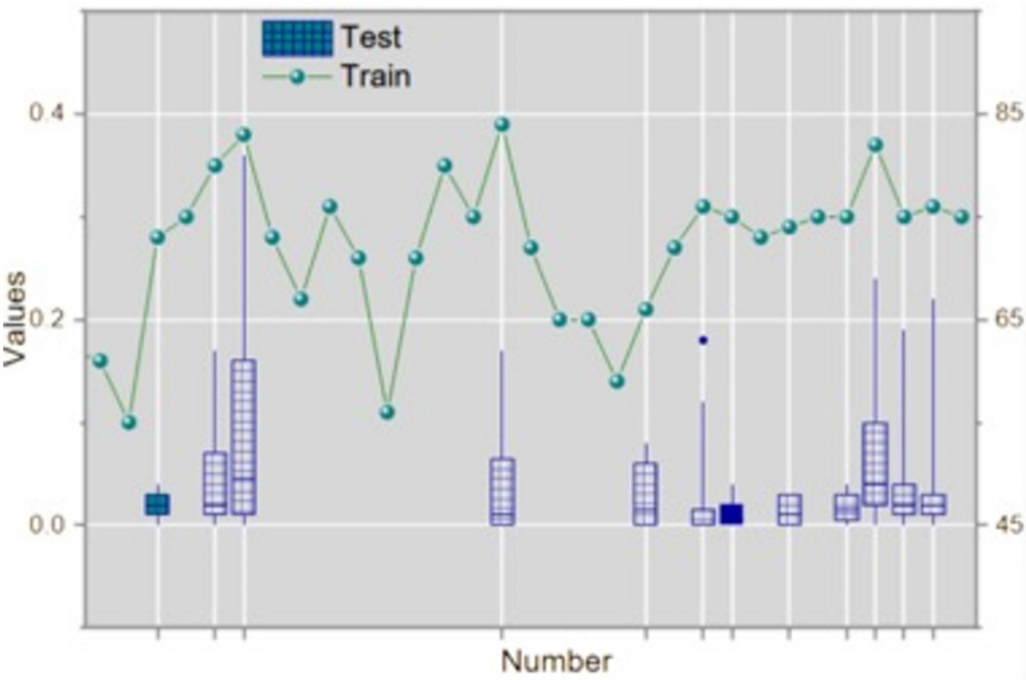
permits (Santana et al., 2017). As another example, some commercial banks granted loans to enterprises that had received multiple environmental protection penalties or were included in the list of non-performing environmental protection enterprises and provided mergers-and-acquisitions (M&A) loans to undocumented cross-border M&A businesses, in serious violation of regulatory policy requirements. There were also individual projects with certificates of construction progress issued by the borrowers, without seeing the verification reports of construction progress confirmed by third-party supervisory agencies or stating whether the project progress matched the amount already invested. Worse still, there was an inadequate review of the availability and commissioning of the project capital, with supporting documents replaced by a statement of circumstances.

The *model training process*, which can also be called the machine learning process, is the process of continuously constructing test set data, performing operations using the current model, and continuously improving the model. In the model training process, it is more intuitive to discover the problems with the initial model weights assigned, to discover the changes in the role of indicators in the model, to develop a model adjustment strategy based on the observed results, and to adjust the indicators and weights of the initial model continuously. This practice-based exploration process is very different from the traditional exploration approach. Traditional empirical analysis emphasizes theoretical guidance in the analysis process and discusses the theoretical rationale for each step in the process. The model training process, on the other hand, constructs the model entirely from the data perspective and can be used as a mature model regardless of whether the model structure conforms to the theory if it can accurately distinguish the identification of the target variables. Therefore, the most significant advantage of the model training process is that it discards the subjective influence and explores the correlation between the indicators from an objective perspective, regardless of whether the data are in line with the law or not, if the conclusions extracted from the data are real and usable, which is in line with the practical standard of “existence is reasonable,” as shown in Table 1.

Although the amount of loans for micro-enterprises is smaller, all need to be the same as medium and large enterprises to carry out credit investigations of corporate assets. Due to the special characteristics of micro-enterprises, their corporate public information and the actual corporate information have a large difference. Business conditions change faster, while the professional managers themselves have to deal with a lot of new customers, so they need more and more mature credit specialists to assist, which will help reduce the risk of ICBC loan funds (Kong et al., 2019). Compared to the loan business of medium and large enterprises, ICBC needs to do a lot of credit investigation work and post-management work before issuing loans, which largely increases the workload of staff but also requires managers more field visits and check the authenticity of the information, only to determine the authenticity of the information to reduce the risk of loans. The amount of information investigation is very large, and the dramatic increase in workload, so that the original scarcity of managers creates more pressure, as shown in Figure 4.

Observing the data in Figure 4, we can find that the scores of all indicators in the test set of the model are all lower than those in the training set. Since the wind control model built by the logistic regression algorithm is the result of continuous training of the training set data, it is easy to over-fit the data during the training process, resulting in a lower score when the model makes predictions on the test set data (Eckhoff & Wagner, 2017; Ma & Lv, 2019). To further improve the wind control model based on the logistic regression algorithm, we next tune the super-parameter of the model. When

Figure 4. Initial logistic regression algorithm performance metric graph



credit money is no longer used as a paper entity but as dematerialized deposit money that takes the form of a bookkeeping symbol on the bank's books, the money flow becomes merely an adjustment between accounts. Thus, credit money, also known as book-entry money, acts as a value measure and a medium of exchange, i.e., a medium for efficient transactions under a broad consensus, with the attributes of ease of calculation, value consensus, high yield, and low exchange costs. As a system of credit, money is not essentially a commodity but a bookkeeping method or a bookkeeping tool, and the credit-money system is a system of bank books (Grace & Williams, 2016; Du et al., 2021).

RESULTS

Credit granting review and risk management are crucial parts of the lending business for banks and financial institutions. The process includes application collection and preliminary examination, which is carried out by verifying the authenticity and completeness of the applicant's identity, income, assets, and other information, as well as conducting a credit assessment, which takes into account factors such as the customer's credit history, repayment ability, and debt burden. Subsequently, the bank conducts loan approval, decides whether to approve the loan application, and determines the loan amount, interest rate, repayment period, and other conditions. If the customer provides collateral, the bank may evaluate it. Once the loan application is approved, the bank and the customer will sign a loan contract to specify the rights and responsibilities of both parties. Finally, the bank will disburse the loan to the customer, monitor the repayment status of the customer, and take necessary measures to deal with late payment or default. This series of processes ensures that the loan disbursement process is compliant and risk-controllable.

Results of Improved Machine Learning Algorithms for Lending Decisions

The idea of model fusion is to fuse multiple base models into a new model with better performance than a single base model. The previously mentioned integrated models, random forest and lightGBM, are also essentially fusion models, resulting from the fusion of multiple tree models. In the previous experiments, we tried to apply logistic regression algorithm, support vector machine algorithm based on different kernel functions, decision tree algorithm, random forest algorithm, and light algorithm to Internet financial risk control model, according to the model scoring results of different algorithms, we will pick out the logistic regression algorithm with better model performance, support vector machine algorithm based on the linear kernel, support vector machine algorithm based on a Gaussian kernel vector machine algorithm, random forest algorithm, and lightGBM algorithm for model fusion. We first picked the best-performing machine learning algorithm among the above models as the target classification model and then used the probabilities predicted by the selected machine learning algorithm models, like the new data features for the target classification model to be trained to train the new fusion model, which is the fusion method of stacking mentioned in the previous section. This paper presents the best-performing logistic regression algorithm as the target classification model and a linear kernel-based support vector machine algorithm, Gaussian kernel-based support vector machine algorithm, random forest algorithm, and light algorithm for stacking fusion.

The statistical results are shown in Figure 5 by measuring the performance of the constructed fusion model on the test set. Figure 5 shows that the value of the fused model is already close to 0.8, which is the highest score among all fused base models, and the effect of model fusion can be seen. Since the fusion model itself is the result of stacking fusion with tuned base models, there is no further tuning process as there is no tuned super-parameter itself. To further understand the effectiveness of different types of machine learning algorithms and the performance differences of fusion models on Internet financial risk control models, we will next explore the results through comparative experiments.

To understand the application effects and performance differences of different types of machine learning algorithms and fusion models on Internet financial risk control models, this paper uses single machine learning algorithms (such as the logistic regression algorithm, support vector machine algorithm based on different kernel functions, and decision tree algorithm), integrated machine learning algorithms (e.g., the random forest algorithm and lightGBM algorithm) and fused machine learning algorithms for comparison experiments, where the support vector machine algorithms based on different kernel functions are represented by the optimal linear support vector machine algorithm, and the comparison results are shown in Table 2.

From Table 2, it can be found that logistic regression among the single models has the highest AUC value score, and, among all the models, the fusion model has the highest AUC value score of 0.7838. Most of the machine learning models have an AUC value of around 0.77, and the accuracy of prediction scores high, around 0.78, which shows the feasibility and effectiveness of machine learning algorithms applied to Internet financial risk control models. To consider the performance of each model comprehensively, we visualize the table as shown in Figure 6.

Figure 6 shows the overall performance of the model is best for the fusion model, followed by the logistic regression model and the lightGBM model, which are neck and neck, and the worst is the decision tree model. In addition, random forest and lightGBM are essentially a fusion of tree models, and the comparison with a single decision tree model further validates the effectiveness of model fusion in improving model performance. In summary, we understand the difference and performance difference between different types of machine learning algorithms applied to the wind control model by performing wind control model construction and results from analysis for single machine learning algorithms (i.e., logistic regression, support vector machines, and decision tree algorithms) and integrated machine learning algorithms respectively. Finally, the performance of the models is further enhanced by stacking fusion, and the individual models are compared and studied, thus laying the theoretical and practical foundation for the following research conclusions.

Table 2. Performance comparison of the models

Name	AUC value	F_1-score	Accuracy	Recall rate
Logistic regression	12.78.	5.73.	7.03.	5.5.
Linear support vector machine	10.29.	3.39.	4.33.	4.45.
Decision tree	1.81.	1.73.	5.45.	14.42.
Random forest	9.43.	13.05.	8.22.	10.15.
Light GBM	4.21.	10.87.	7.13.	12.94.
Stacking integration	6.88.	1.46.	7.18.	9.57.

Results of Experiments on Bank Finance-Type Lending Decision Mechanisms

For the same processed financial data, the model performance of support vector machine algorithms with different kernel functions varies greatly. The best model performance of linear support vector machine AUC value can reach 0.779, close to 0.78, but the worst performance of polynomial support vector machine algorithm, the AUC value is only 0.734, less than 0.75, and the other two kernel functions of support vector machine models are only 0.75 or so. The types of data suitable for processing are different for different kernel functions, whereas linear support vector machines are more suitable for easily linearly separable data. Also, comparing the logistic regression model and the random forest model, random forest is an advanced tree model, while logistic regression is just

Figure 5. Fusion model performance metric graph

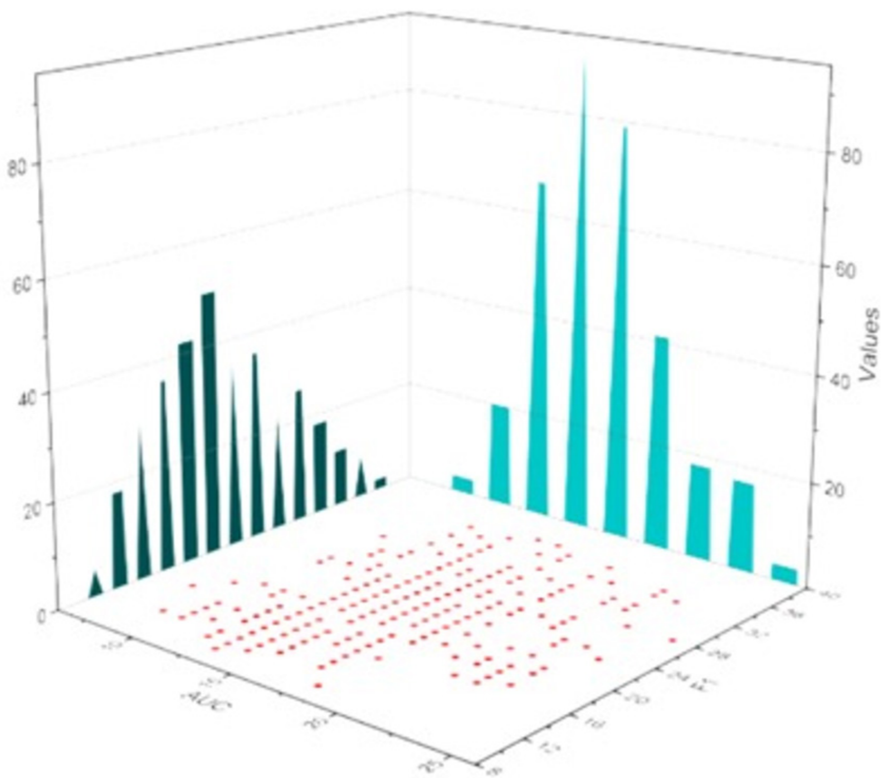
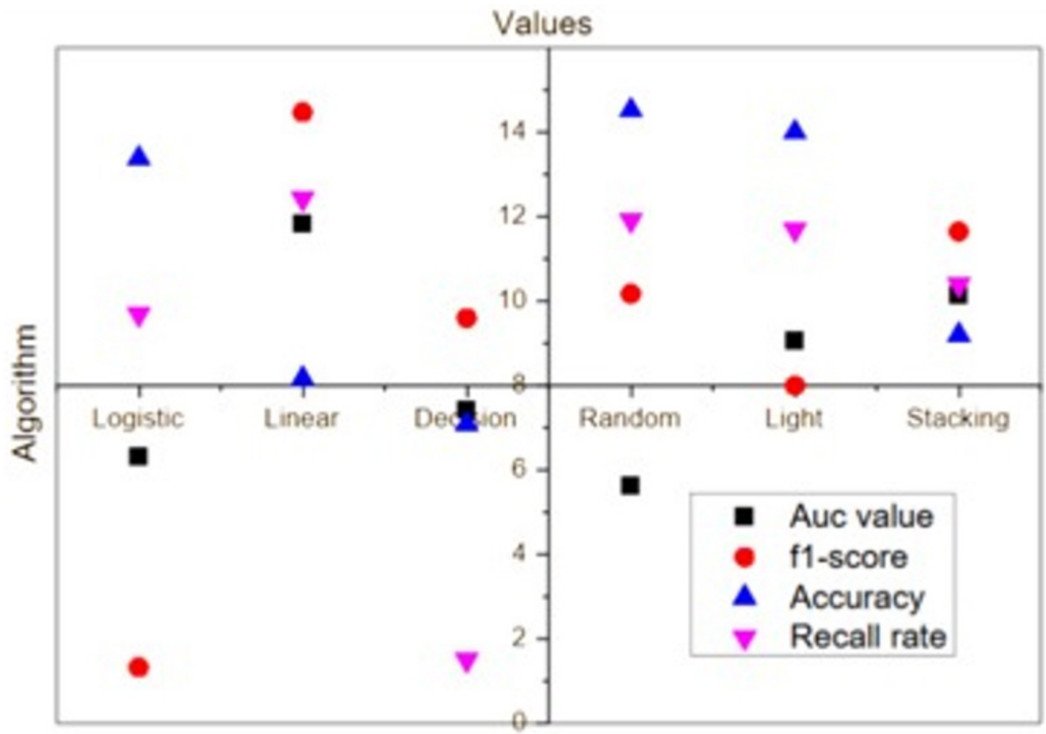


Figure 6. Comparison of performance measures across models

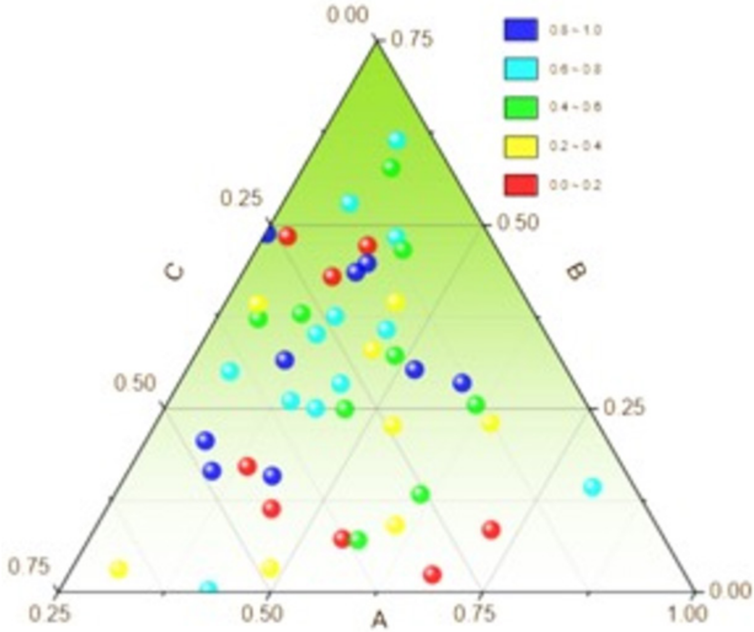


an ordinary linear model, but the model performance of the logistic regression model is even better than random forest. Throughout the comparison of models, both the logistic regression algorithm and linear support vector machine algorithm, which are also linear models have better model performance. The more advanced the model is, the better it is, and simple linear models sometimes perform better. Therefore, machine learning algorithms that are suitable for the data type can effectively improve the model performance, and a lot of time and effort can often be saved if the data type is understood first when the model is selected, as shown in Figure 7.

In general, when a model is constructed, the generalization ability of the model needs to be improved, i.e., the gap between the model's performance on the validation set and its performance on the training and test sets is narrowed by one of the following methods: retraining the model by (a) increasing the sample size; (b) adding new variables, i.e., increasing the amount of information in the variables entering the model to improve the overall predictive power of the model; or (c) examining the population characteristics of the sample to see if the population characteristics across time whether the population characteristics of the validation set are significantly different from those of the training set. Since the modeling sample can no longer be increased, finally this study used the original bottom search data of users included in the company's search engine to join the model, to improve the generalization ability of the model. The data of the original bottom search terms are of high dimensionality but sparse, so they cannot be directly added to the logistic regression as variables.

In a general sense, the user's historical search behavior can reflect the user's personal preferences, personality traits, and other aspects of information to a certain extent, and variables such as the company's portrait of the user and the user's high-risk search terms are also derived variables made based on these primitive search terms. Based on the high dimensionality and sparse data of these primitive underlying search terms, this paper decides to model these primitive underlying search terms

Figure 7. Relationship between user churn rate and bank loan rates for different classes



using the GBDT model and adds the resulting user default probability value to the model as a new variable in the initial version of the logistic country regression model, to improve the generalization ability of the model. Commercial banks can fit the relationship between different levels of customer churn and bank lending rates by establishing functional equations, and if internal customer reputation ratings are missing, they can also use historical data to model multi-classification problems based on unsupervised learning such as anomaly detection algorithms for the predictive variables of a firm's reputation ratings. The SME lending business of commercial banks is completed under the model construction of the system, which involves more enterprise data that requires continuous supervision before, during, and after lending, and further monitoring of risks under the guidance of interest rate marketization.

CONCLUSION

Through the analysis of the experimental results, it is found that the fusion model shows the best performance among the Internet financial risk control models, followed by the logistic regression and lightGBM models, while the decision tree model performs poorly. In particular, in the experiment of bank finance-type loan decision mechanism, the linear support vector machine algorithm performs the best when dealing with data that are easily linearly separated, while the logistic regression model performs even better than the random forest model. When choosing a model, a suitable choice should be made based on the complexity of data types and features, and sometimes a simple linear model can achieve better results. In addition, the generalization ability of the model is improved by introducing the original underlying user search data, which further strengthens the predictive ability of the model. Commercial banks can fit the relationship between customer churn and bank loan interest rates in SME lending business by establishing functional equations, while using historical data for unsupervised learning in order to improve the customer reputation rating system and continuously supervise the risks to optimize the lending decision-making mechanism further.

DATA AVAILABILITY

The figures used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

No funding was received for this work.

PROCESS DATES

Received: April 8, 2024, Revision: April 19, 2024, Accepted: April 22, 2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Anna Quaranta (Italy, annagrazia.quaranta@intro.college)

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74. 10.1186/s40537-021-00444-833816053

Angelidou, M. (2017). The role of smart city characteristics in the plans of fifteen cities. *Journal of Urban Technology*, *24*(4), 3–28. 10.1080/10630732.2017.1348880

Aphale, A. S., & Shinde, S. R. (2020). Predict loan approval in banking system machine learning approach for cooperative banks loan approval. *International Journal of Engineering Trends and Applications*, *9*(8).

Bazarbash, M. (2019). *Fintech in financial inclusion: machine learning applications in assessing credit risk*. International Monetary Fund.

Caird, S. P., & Hallett, S. H. (2019). Towards evaluation design for smart city development. *Journal of Urban Design*, *24*(2), 188–209. 10.1080/13574809.2018.1469402

Chen, M., Yang, F., & Choi, Y. (2021). Are credit-based Internet consumer finance platforms sustainable? A study on continuous use intention of Chinese users. *Sustainability (Basel)*, *13*(24), 13629. 10.3390/su132413629

Du, G., Liu, Z., & Lu, H. (2021). Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment. *Journal of Computational and Applied Mathematics*, *386*, 113260. 10.1016/j.cam.2020.113260

Eckhoff, D., & Wagner, I. (2017). Privacy in the smart city: Applications, technologies, challenges, and solutions. *IEEE Communications Surveys and Tutorials*, *20*(1), 489–516. 10.1109/COMST.2017.2748998

Gharaibeh, A., Salahuddin, M. A., Hussini, S. J., Khreishah, A., Khalil, I., Guizani, M., & Al-Fuqaha, A. (2017). Smart cities: A survey on data management, security, and enabling technologies. *IEEE Communications Surveys and Tutorials*, *19*(4), 2456–2501. 10.1109/COMST.2017.2736886

Grace, A. M., & Williams, S. O. (2016). Comparative analysis of neural network and fuzzy logic techniques in credit risk evaluation. *International Journal of Intelligent Information Technologies*, *12*(1), 47–62. 10.4018/IJIIT.2016010103

Kakderi, C., Komninos, N., & Tsarchopoulos, P. (2019). Smart cities and cloud computing: Lessons from the STORM CLOUDS experiment. *Journal of Smart Cities*, *1*(2), 4–13.

Komninos, N., Kakderi, C., Panori, A., & Tsarchopoulos, P. (2019). Smart city planning from an evolutionary perspective. *Journal of Urban Technology*, *26*(2), 3–20. 10.1080/10630732.2018.1485368

Kong, X., Liu, X., Jedari, B., Li, M., Wan, L., & Xia, F. (2019). Mobile crowdsourcing in smart cities: Technologies, applications, and future challenges. *IEEE Internet of Things Journal*, *6*(5), 8095–8113. 10.1109/JIOT.2019.2921879

Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, *63*(2), 157–170. 10.1016/j.bushor.2019.10.005

Ma, X., & Lv, S. (2019). Financial credit risk prediction in Internet finance driven by machine learning. *Neural Computing & Applications*, *31*(12), 8359–8367. 10.1007/s00521-018-3963-6

Mehmood, Y., Ahmad, F., Yaqoob, I., Adnane, A., Imran, M., & Guizani, S. (2017). Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine*, *55*(9), 16–24. 10.1109/MCOM.2017.1600514

Mohanty, S. P., Choppali, U., & Kougianos, E. (2016). Everything you wanted to know about smart cities: The Internet of things is the backbone. *IEEE Consumer Electronics Magazine*, *5*(3), 60–70. 10.1109/MCE.2016.2556879

Orlova, E. V. (2020). Decision-making techniques for credit resource management using machine learning and optimization. *Information (Basel)*, *11*(3), 144. 10.3390/info11030144

Pahlevan Sharif, S., & Naghavi, N. (2020). Family financial socialization, financial information seeking behavior and financial literacy among youth. *Asia-Pacific Journal of Business Administration*, *12*(2), 163–181. 10.1108/APJBA-09-2019-0196

Panori, A., González-Quel, A., Tavares, M., Simitopoulos, D., & Arroyo, J. (2019). Migration of applications to the cloud: A user-driven approach. *Journal of Smart Cities*, *2*(1), 16–27.

Rana, N. P., Luthra, S., Mangla, S. K., Islam, R., Roderick, S., & Dwivedi, Y. K. (2019). Barriers to the development of smart cities in Indian context. *Information Systems Frontiers*, *21*(3), 503–525. 10.1007/s10796-018-9873-4

Santana, E. F. Z., Chaves, A. P., Gerosa, M. A., Kon, F., & Milojicic, D. S. (2017). Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture. *ACM Computing Surveys*, *50*(6), 1–37. 10.1145/3124391

Sima, V., Gheorghe, I. G., Subić, J., & Nancu, D. (2020). Influences of the industry 4.0 revolution on the human capital development and consumer behavior: A systematic review. *Sustainability (Basel)*, *12*(10), 4035. 10.3390/su12104035

Trencher, G., & Karvonen, A. (2020). Stretching “smart”: Advancing health and well-being through the smart city agenda. In *Smart and sustainable cities?* (pp. 54–71). Routledge.

Tsarchopoulos, P., Komninos, N., & Kakderi, C. (2017). Accelerating the uptake of smart city applications through cloud computing. *International Journal of Economics and Management Engineering*, *11*(1), 129–138.

Turkson, R. E., Baagyere, E. Y., & Wenya, G. E. (2016, September). A machine learning approach for predicting bank credit worthiness. In *2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR)* (pp. 1-7). IEEE. 10.1109/ICAIPR.2016.7585216

Wu, Q., & Zhang, B. (2020, December). The influence of Internet finance on the traditional business of commercial banks. *Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologie*s, 570–575. 10.1145/3444370.3444629

Xiao, J. J., & Tao, C. (2021). Consumer finance/household finance: the definition and scope. *China Finance Review International*, 1^1(1), 1–25.

Xie, J., Tang, H., Huang, T., Yu, F. R., Xie, R., Liu, J., & Liu, Y. (2019). A survey of blockchain technology applied to smart cities: Research issues and challenges. *IEEE Communications Surveys and Tutorials*, *21*(3), 2794–2830. 10.1109/COMST.2019.2899617

Zheng, C. (2022). Complex network propagation effect based on SIRS model and research on the necessity of smart city credit system construction. *Alexandria Engineering Journal*, *61*(1), 403–418. 10.1016/j.aej.2021.06.004

Ziemba, P., Becker, J., Becker, A., Radomska-Zalas, A., Pawluk, M., & Wierzba, D. (2021). Credit decision support based on real set of cash loans using integrated machine learning algorithms. *Electronics (Basel)*, *10*(17), 2099. 10.3390/electronics10172099

Ziemba, P., Becker, J., Becker, A., Radomska-Zalas, A., Pawluk, M., & Wierzba, D. (2021). Credit decision support based on real set of cash loans using integrated machine learning algorithms. *Electronics (Basel)*, *10*(17), 2099. 10.3390/electronics10172099