Risk Analysis of Campus Loan based on LGBM from the Perspective of Soft and Hard Information

Haixia Lv

Shandong University of Finance and Economics, Jinan, Shandong, China

xzdjl@126.com

Keywords: Campus Loan, Soft and Hard Information, Lgbm, Binary Classification, Analysis and Prediction

Abstract: Campus loan refers to the installment payment or loan obtained by college students through the online loan platform, which solves the problem of shortage of funds for some students. However, due to the non-standard development of "campus loan" and its increasingly hidden nature, malignant events still occur frequently, which has a very bad impact on the school and society. From two aspects of soft and hard information on college students' campus loans to investigate analysis, firstly the questionnaire survey was distributed to 2735 college students randomly selected in Shan Dong provincial universities as soft information, and analyze the campus loan according to the results of the study some relevant situations of existence and development of colleges and universities in ShanDong province. Then from the point of view of hard information, this article takes the UCI default of credit card client data sets, using histogram based decision tree algorithm LGBM(Light Gradient Boosting Machine) algorithm of decision tree, training and making prediction of data set. Finally, through the analysis of the essence of campus loan from two aspects of hard and soft information, and from the perspective of the synergistic effect of government, university, enterprise and family, puts forward countermeasures to prevent the risk of campus loan.

1. Introduction

"Campus loan" is a kind of credit loan with college students as the lending objects. Its essence is an online credit consumption loan with college students as the lending objects. It is the product of the great development of Internet finance. As a new lending method, it has the advantages of low application threshold, fast approval speed, and no need for financial guarantees. It meets the strong demand of college students for credit consumption and the urgent desire to pursue "financial independence". However, in the process of development, there have also been some loan platforms that engage in false advertising, information concealment, credit defaults, high borrowing risks, setting up illegal lending traps, inducing consumption, high compound interest rates, violent debt collection, and other phenomena, seriously disrupting financial regulatory order and infringing on the personal and property safety of loan students. Some students are even forced to commit illegal crimes or commit suicide due to campus loans. Due to legal gaps, regulatory deficiencies, and lagging control, the drawbacks of "campus loans" have gradually emerged, leading to the phenomenon of "campus loans" and the emergence of social problems.

In response to a series of problems caused by "campus loans," the Ministry of Education and the China Banking Regulatory Commission jointly issued a notice in April 2016 on strengthening the prevention and education guidance of risks in non-performing online loans on campus. The notice clearly stipulates that no online loan institution is allowed to issue loans to college students. Other normative documents such as the "Notice on Further Strengthening the Rectification of Campus Online Loans" and the "Notice of the General Office of the Ministry of Education on Carrying out Risk Warning Education and Related Work on Campus Non performing Online Loans" have also proposed policies of "stopping, moving, rectifying, teaching, and guiding", and explicitly ordered the temporary suspension of "campus loans". However, due to the limited social experience of college students, weak prevention awareness, poor self-control, and susceptibility to impulsive consumption,

  DOI: 10.25236/meeit.2024.028

and the increasingly hidden characteristics of "campus loans", new forms such as "entrepreneurial loans", "training loans", and "exam loans" have emerged. The phenomenon of "campus loans" cannot be completely eradicated, and the "campus loan" incident has not been effectively curbed. A series of tragedies caused by "campus loans" have emerged. Multiple provinces across the country have repeatedly staged: "College students are trapped in routine loans to buy mobile phones" "After borrowing 6000 yuan for a month, I was forced to repay 120000 yuan." "A college student borrowed 3000 yuan online to start a micro business, but was scammed out of 100000 yuan to repay old debts by borrowing new ones.". Therefore, in order to further understand the current phenomenon of campus loans, explore the factors that have an impact on campus loan defaults, and study and analyze the existing ways of the current phenomenon of "campus loans" in universities, this article will draw on the concepts of soft and hard information in the field of financial management, comprehensively analyze and predict, explore its essence and response strategies.

2. Related work

In the field of finance and economics, we often come into contact with two types of information: soft information and hard information [1]. Specifically, soft information is numerical data that cannot be measured by clear scales or specific indicators, which includes non-financial information such as social capital, status, interpersonal relationships, and personal reputation; on the other hand, information that can be represented by precise numerical values of a certain scale is called hard information, which includes financial information such as age, gender, annual income, etc.

In a broad sense, the main differences between them are as follows: in terms of standardization, the hard information format is standardized and obtained based on established rules, while the soft information format is not determined; in terms of timeliness, soft information is randomly generated and have unpredictability, while the acquisition cycle of hard information has regularity; in terms of accuracy, hard information is more precise, while soft information does not have an objective evaluation standard. However, there is a corresponding connection between them, which is not completely isolated. Hard information is the foundation for the generation of soft information and the com-bination of hard and soft information can be integrated into a continuous and unified whole. Meanwhile, there is a natural transitional relationship between soft and hard information. When subjective factors account for a large proportion of the total information, hard information becomes soft information.

Currently, most of the existing research in China unilaterally studies the impact of hard information on loan defaults. Among them, Angelini et al. [2] propose earlier to apply neural network technology to credit risk prediction research. Tang et al. [3] use SAS statistical software to analyze financial data using full samples, applied Bayes discriminant principle, introduced misjudgment costs and prior probabilities, and constructed a concise default discrimination model. Han et al. [4] propose a credit risk portfolio prediction and evaluation method based on fuzzy neural networks, which comprehensively considers the impact of financial and non-financial factors on credit risk. Zhang et al. [5] propose a support vector machine model to construct a technology credit default prediction model for small and medium-sized enterprises by constructing different input variable evaluation index systems. Wang et al. [6] use Logistic, XGBoost and NN to predict default risk for online loan borrower loan risk data. Zhou [7] proposed an improved weighted parallel balanced random forest algorithm (WPBRF) for imbalanced loan default data. Barbaglia [8] analyze at the regional level, modeling the occurrence of default as a function of borrower characteristics, loan specific variables, and a set of local economic conditions. Liang [9] use LSTM to model and predict the monthly default rate of P2P platforms. Arian [10] propose a method based on Gaussian Mixture Model (GMM) to predict the default probability of individual loan applicants by clustering similar customers together. Odegua [11] use XGBoost to predict the loan risk of applicants based on loan data, taking into account both loan application and applicant demographic datasets. Mi [12] use the Multinomial Lasso logistic model to identify the key factors affecting platform default and predict the probability of default. Ma et al. [13] take the Renren Loan platform as the research object, apply feature engineering technology, and apply the CatBoost algorithm to construct a P2P default prediction model, and comprehensively analyze the factors affecting default. Chen et al. [14] conduct quantitative research using logistic regression models and stepwise regression models based on questionnaire surveys. However, considering the impact of non-financial information on customer credit default or loan default, relevant research clearly lacks attention to non-financial information. Chen et al. [15] conduct a comparative study on online lending practices in the United States and China, and find that two types of credit information, "hard" and "soft", may have a profound impact on loan outcomes in both countries. Guo et al. [16] use loan default data from a domestic commercial bank for small and medium-sized enterprises, and use random effects logistic models to empirically analyze the influencing factors of loan default risk for non-listed small and medium-sized enterprises. Jiang et al. [17] use topic models to extract and quantify relevant variables from text soft information, and then analyze the impact of different soft information variables on loan defaults.

This article analyzes campus loans from two aspects. On the one hand, it uses soft information to collect and analyze student participation in campus loans through a campus survey questionnaire. On the other hand, based on the information samples of online user defaults, combined with the LGBM binary classification algorithm, the samples are predicted.

3. Soft information investigation and analysis

3.1. Investigation and Analysis of the Current Situation of "Campus Loans"

In the form of a questionnaire survey, 2735 students from the 2014-2017 cohort were randomly selected to participate in the survey. The questionnaire was distributed and filled out using the "QuestionStar" mobile app. A total of 2735 valid respondents were received, and data statistical analysis was conducted. The investigation results and cause analysis are as follows:

Participation and Analysis Survey of "Campus Loan". The results showed that 28.3% of students had used similar products such as Ant Huabei Jiebei, JD Campus Baitiao, Qu Fen, Ren Fen, and Famous School Loan, indicating a high level of participation in campus loans among students from financial and economic colleges. In the "Have your friends around you had any experience using campus loans" section (Fig.1), 39.23% said "never", 29.47% said "very rarely", and 19.71% said "yes, installment shopping is still popular". These data indirectly indicate that the situation of campus loans is quite common. 54.99% of the students who participated in the survey questionnaire were freshmen, indicating that lower grade students have a higher level of participation due to their light academic burden, strong curiosity about new things, lack of self judgment, and lack of identification and resistance to campus loans.

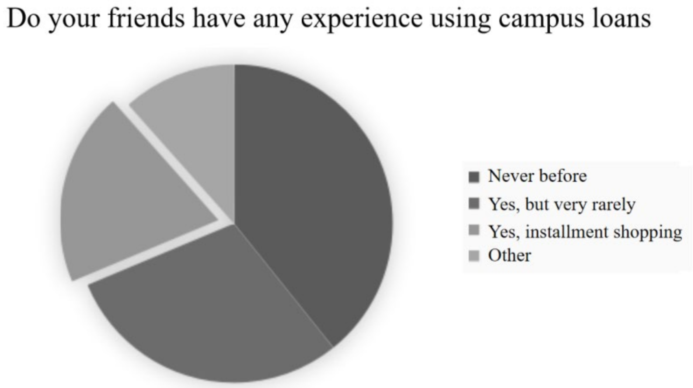


Figure 1 Participation in Campus Loans.

Cognitive situation and interest rate of "campus loan" products. In the questionnaire, the campus loan products that students are relatively familiar with were listed, and classification options were designed. When choosing the "campus online loan platform I've heard of", the proportion of Alipay Ant Flower Pai is as high as 89.4%, and the proportion of JD Campus Baitiao is 33.71%, which indicates that finance and economics students have some contact with many products of "campus loan". Among them, Alipay Ant Flower Pai, JD Campus Baitiao and other products relying on Alipay, JD and other large-scale shopping platforms are more popular with finance and economics students.

Most of these campus loan products are consumption installments, and do not participate in credit cashing, which indicates that finance and economics students value the safety of campus loan products. However, regarding the calculation of interest rates for campus loans (Fig.2), 53.6% of students have little understanding, while only 6.58% of students indicate that they have calculated and are quite familiar with it. The above data indicates that a considerable number of students do not have sufficient understanding of the interest rate mechanism of campus loan products, and do not have a sufficient understanding of the risk of future defaults, which poses significant risks and hidden dangers.

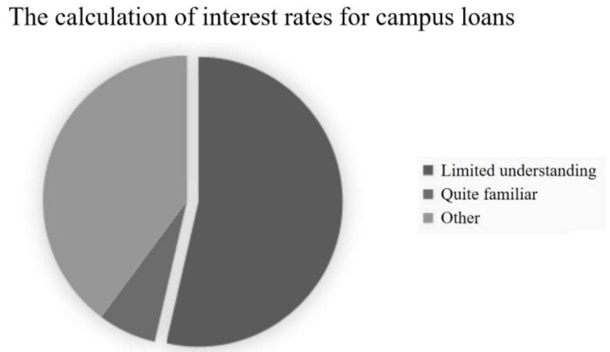


Figure 2 Cognitive situation of campus loan interest rates.

Analysis of Student Consumption and Usage. According to the survey results, students spend 40.44% of their monthly living expenses in the range of 500-1000 yuan, 53.42% in the range of 1000-2000 yuan, and a very low proportion exceeding 2000 yuan. In the item of whether the actual monthly consumption exceeds the cost of living, 48.1% of students occasionally exceed it, and 3.4% of students often consume ahead of schedule. It indicates that the current economic level of students is relatively affluent in daily life, and there is a phenomenon of excessive consumption. Only less than one-third of students are able to plan their living expenses reasonably.

When faced with the problem of "if there is excessive consumption but unable to repay it, __", 1.76% of students choose to borrow new and old loans, and look for other loan platforms to borrow, indicating that after excessive consumption, college students mainly ask their parents for money or make money on their own. Only a few people choose "campus loans", but there are also many platforms that use the needs of students for work, part-time jobs, and other marketing. If students have excessive consumption, they also have to bear certain potential risks.

Figure 3 shows the results of the question "the purpose of student consumption" (multiple choice question), 46.54% of students choose electronic products, 52.5% of students choose clothing and beauty, 41.13% of students choose tourism consumption, 26.76% of students choose socializing (dating, dining, etc.), and 34.63% of students choose entrepreneurship. The above data is in line with the consumption identity and practical needs of students, and the proportion of entrepreneurial students reflects the good momentum of student entrepreneurship phenomenon.

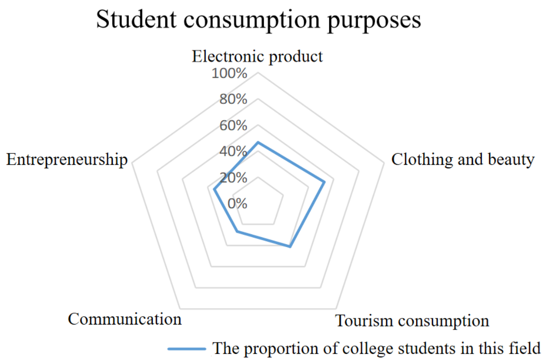


Figure 3 The proportion of student consumption purposes.

Attitude and understanding towards campus loans. The results show that students hold a neutral attitude towards campus loans, and 56.1% of them believe that the economic situation allows them to use them; 33.97% of students oppose campus loans and strongly oppose them, indicating a high acceptance of campus loans. As for the nature of campus loans, 45.27% of students think campus loans are usurious loans, 24.86% think campus loans are ordinary loans, 20.37% think campus loans belong to Internet finance, and 9.51% think campus loans are innovative financial products, which shows that college students have diverse understanding of the nature of campus loans, but lack of awareness of the risks involved.

Risks, problems, and measures of campus loans. In the survey, "What do you think are the main risks of campus loans?" (multiple choice question), 59.6% chose low entry barriers, less strict review, and low reliability of agents. 57.99% chose high interest rates, management fees, and service fees, and many other charging items. 64.79% chose to disclose personal information. 50.35% chose chaotic loan platforms and incomplete laws. The survey results indicate that students have a certain understanding of the risks of campus loans. When faced with the question of "how to view the risks of campus loans", 50.49% of students believe that there is a risk but it can be avoided, and 11.92% of students believe that there is no risk and timely repayment is sufficient. This result indicates that students hold a relatively optimistic attitude towards the risks of campus loans.

In the survey, 77.88% of the students believe that the main problem of campus loan consumption is to encourage some students' improper consumption habits and impulsive consumption, and 27.61% believe that colleges and universities still lack education in Internet finance knowledge and healthy consumption concepts. 70.38% of students believe that college students refuse impulsive consumption and form timely, moderate, and reasonable consumption habits. 63.4% of students believe that online lending platforms should strengthen the approval process, standardize lending, repayment, and collection procedures and processes. 61.06% of students believe that the government should increase monitoring and management of campus loans. 43.36% of students believe that schools should carry out special education to improve students' financial management ability and loan risk awareness. 10.24% of students believe that various sectors of society should widely promote and coordinate with each other.

3.2. Performance and reasons of campus loan risks

Campus loans are usually divided into the following five types: credit services provided by traditional e-commerce platforms such as Taobao and JD.com, cash withdrawals provided by consumer finance companies, P2P loan platforms, high interest loans from private lending institutions and lenders, and campus products provided by banks to college students.

Campus loans often contain fraudulent elements, using the gimmick of "having no physical collateral, having a certificate to make money, fast loan disbursement speed, being able to handle on behalf of others, zero down payment, zero interest, etc." to conceal the detailed information of loans and induce students to borrow. In terms of promotional interest rates, the actual interest rate of campus loans is much higher than the promotional interest rate. Therefore, campus loans are not uncommon on campus under the guise of low interest rates. Campus loans are essentially an invisible high interest loan aimed at students on campus, a P2P network financial form that survives with high transaction fees, intermediary fees, management fees, service fees, breach of contract fines, and other fees.

From the survey questionnaire, it can be seen that college students have a certain understanding of the risks of campus loans, and the reason why the college student group is trapped in the borrowing quagmire is mostly due to irrational consumption concepts and advanced consumption methods. According to Maslow's hierarchy of needs theory, contemporary college students are no longer just satisfied with survival needs, but more importantly, they pursue the quality of life. Many college students pursue branded clothing, bags, high-end electronic products, high-end restaurants, and satisfy their vanity psychology. To form a false need that is not in line with one's own economic ability and contradicts the practicality of reality, in order to curb the harm of college student campus loans, fundamentally speaking, it is necessary to carry out work from the ideological source, eliminate blind consumption, comparison psychology, reject impulsive consumption, and follow the trend of consumption, continuously forming a scientific, rational, civilized, and moderate consumption concept.

Secondly, college students have a low awareness of risk prevention. From the survey questionnaire, it can also be seen that college students generally lack sufficient risk prevention for campus loans, and have a mentality of taking chances. This is related to the immature mentality, lack of social experience, simple thinking, and psychological fragility of the college student group. They often easily believe that false advertising on the surface does not study detailed items, and they do not consider their personal information security issues. Once they are deceived into not repaying on time, online loan companies resort to violent debt collection. Under pressure, college students often "demolish the east wall, pay the west wall", and fall into a vicious cycle of loans, carrying a heavy debt burden.

A questionnaire survey on five aspects of soft information, including participation and analysis of "campus loans", product awareness and interest rate calculation of "campus loans", analysis of student consumption and usage, attitude and understanding of campus loans, risks, problems and measures of campus loans. From the perspectives of both the applicant and the borrower, it can be concluded that, from the perspective of the applicant university students, it is common for them to borrow small loans related to campus loans. They are not very familiar with the interest rate mechanism of campus loans and do not have a clear understanding of the risks of campus loans. They pursue the improvement of their quality of life, especially in daily consumption expenses; From the above analysis, it can be seen that the information inferred from the survey questionnaire options has strong guiding significance in the overall direction, laying the foundation and overall direction for predicting loan information with hard information. We continue to make accurate predictions through hard information and, based on the prediction results, play a preventive role in the possible default situations of college students in the future. This example is used as an analysis of hard information for campus loans [18].

4. Hard Information Analysis

In terms of hard information, UCI's default of credit card client dataset is adopted, and the decision tree algorithm LGBM (Light Gradient Boosting Machine) based on histogram is used to train and predict the dataset.

4.1. Models and Methods

The classification decision tree model is a tree structure that describes the classification of instances. A decision tree consists of nodes and directed edges. There are two types of nodes: internal nodes and leaf nodes. Internal nodes represent a feature or attribute, while leaf nodes represent a class. Using decision tree classification, starting from the root node, test a certain feature of the instance, and assign the instance to its child nodes based on the test results; Each child node corresponds to a value for that feature.

The LGBM algorithm can achieve binary classification by iteratively fitting gradients. The core idea is to iteratively update the initial classifier (weak classifier) on the same training set, forming a stronger final classifier (strong classifier). The algorithm itself is achieved by fitting feature gradients. It sends the residual gradients in each training set as new true values of the samples to the lower level classifier for training, iteratively updating the classifier to make the final decision classifier. This algorithm converts feature values from continuous values to discrete values by establishing a histogram, and splits leaf nodes by traversing all features to find the feature with the highest gain and its partition value.

Firstly, initialize the weak learner using the following formula:

$$f_0(x) = \arg\min_c \sum_{i=1}^{N} L(y_i, c) \quad (1)$$

Iteratively updating the strong learner during the training process:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} \gamma_j I(x \in R_j) \tag{2}$$

To obtain the final classifier F (x).

$$F(x) = f_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J} \gamma_{jm} I(x \in R_{jm}) \tag{3}$$

Where x is the input vector, γ_{jm} is the m-th iteration calculated based on gradient residuals, the best fit value of the jth leaf node.

Compared to other machine learning based classification algorithms, LGBM has the following advantages: firstly, the histogram divides continuous feature values into k discrete feature values, reducing memory consumption and computational costs; Secondly, the optimization algorithm of LGBM adopts a depth limited leaf wise algorithm, which can achieve parallel computing, accelerate the training process, and prevent overfitting; Finally, the construction of the histogram can be accelerated by calculating the difference of the histogram, in addition to relying on the gain.

4.2. Training Algorithm

During training, initialize a weak classifier. By adjusting the decision tree parameters to fit the feature gradient, the optimization strategy for gradient fitting adopts a histogram based leaf wise growth algorithm. By continuously searching for the leaf node with the highest profit after splitting, it is split into two nodes and grown into a decision tree. The use of histograms simplifies the search for gradients and accelerates the calculation of returns. The pseudocode process for training algorithms is as follows:

Algorithm 1 LGBM Classifier training algorithm:

Input: Given l training samples (x_i, y_i), where x_i is the feature vector and y_i is the category label.

Output:

1)Initialize the weak classifier according to formula (1).

2)Calculate gradient residuals for each sample.

3)Establish a histogram H for each sample feature of each leaf node, storing the sum of sample gradients gi and the sum of sample quantities ni.

4)Take the data (xi, yi, gi) as the new sample true value and recalculate the best fit value for each leaf node based on this.

5)Update the classifier according to formula (2).

6)for m to M do repeat steps 2-5.

7)Finally, a strong classifier is obtained, as shown in formula (3).

4.3. Datasets

The UCI database is a shared database of the University of California, Irvine, USA. The data is the 2016 Taiwan customer default payment dataset: default of credit card clients, which is commonly used in data mining to predict default probabilities. The dataset contains up to 30000 customer information, including 24 attribute quantities. The customer's status is divided into two categories: default payment and on-time payment. The number of training and testing sets is the same, with 15000 each.

4.4. Experimental results

In the experiment, the learning rate used in this article is based on the experimental results. After multiple iterations of optimization, the accuracy of the training dataset and validation dataset can be improved to about 80% during training. Fig.4 shows the variation curve of LGBM accuracy during the training process.

This article compares LGBM with Logistics, DTC, XGB, SVM and Adaboost. As shown in Table 1, the accuracy and running time of the predictions of different models mentioned above are compared. In order to better compare running time and accuracy, this article uses cross validation to evaluate all models and find the optimal parameter settings. It can be seen that under the same settings, the LGBM method has higher accuracy than other methods, and the time spent is also less than most second-class classification methods, which is within an acceptable range, as shown in Table 1.

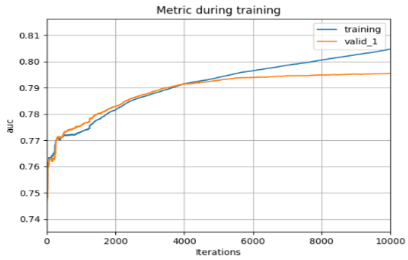
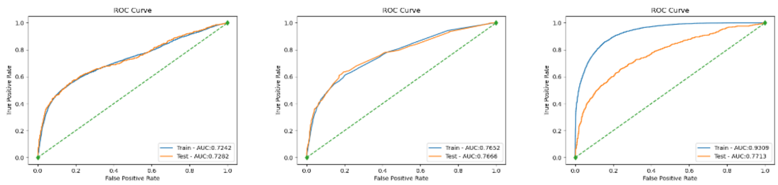


Figure 4 Accuracy variation curve during LGBM training process.

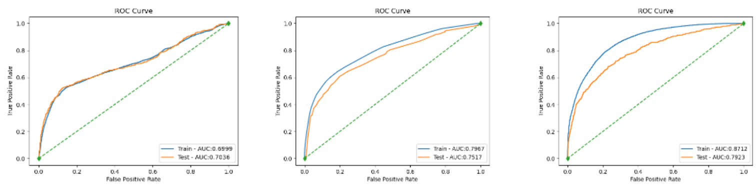
Table 1 Comparison of accuracy and running time of different binary algorithms on the testset.

method	Accuracy	Running time(s)
Logistics	0.809	1.80
DCT	0.813	2.51
XGB	0.819	29.29
SVM	0.815	1456.38
Adaboost	0.814	4.08
LGBM	0.825	3.03

When the distribution of positive and negative samples changes, the shape of the ROC change curve can remain basically unchanged. Therefore, this evaluation indicator can reduce the interference caused by different test sets and more objectively measure the performance of the model itself. Fig.5 shows the ROC variation curves of different models predicting the training and validation datasets during the training process. The AUC (Area under roc Curve) marked in the figure refers to the size of the area under the ROC curve. The larger the value, the better the performance of the model. As shown in Fig.5, LGBM has the highest AUC value, therefore its model performance is the best compared to the other five methods.



a. Logistics b. DCT c. XGB



d. SVM e. Adaboost f. LGBM

Figure 5 ROC variation curves of different models.

During testing, as accuracy increases, recall decreases, and vice versa. F1 score, on the other hand, is a harmonic average of accuracy and recall. Table 2 compares the accuracy, recall, and F1 score of the six models mentioned above. The results show that the LGBM model has a high accuracy, a high F1 score, and good model performance.

4.5. Discussion of experimental results

The LGBM algorithm has higher accuracy compared to other classical algorithms, and the prediction time is also within an acceptable range. This article uses online loan information to predict whether borrowers will default, and achieves good prediction results, fully demonstrating the importance of hard information. The experimental results demonstrate the necessity of reviewing student application information for campus loans, the feasibility of predicting the credit level of college students through application information, and provide relevant technical support for predicting the possibility of college students defaulting in the future.

Table 2 Comparison of Precision, F1-score and Recall of Different Algorithms on the testset.

method	Precision	Recall	F1-score
Logistics	0.750	0.226	0.347
DCT	0.747	0.280	0.280
XGB	0.662	0.366	0.471
SVM	0.700	0.321	0.440
Adaboost	0.699	0.322	0.441
LGBM	0.752	0.350	0.474

5. Risk prevention and governance strategies for campus loans

The investigation and analysis of soft and hard information indicate that in order to prevent, guide, and solve the problem of campus online loans, government departments, online loan companies, schools, families, and students need to work together to form a joint force, face together, propose feasible risk prevention measures, jointly build a comprehensive risk prevention mechanism, establish a countermeasures system for campus loan risk prevention, help college students establish a healthy and reasonable consumption concept and values, cultivate and establish a good credit awareness of college students, and promote college students to face life and learning with a more sunny attitude.

5.1. Improve the market supervision system and form a joint effort between the government and universities

Establish a multi departmental control mechanism to control and prevent risks from the source. At present, domestic universities have not established a comprehensive credit evaluation system for college students, and their credit awareness is seriously lacking. In this regard, schools and the government should work together to establish a unified credit reporting system platform, promote information sharing and data exchange, record the reception credit information of college students, and avoid the problem of overdue loans for college students. Actively promote the construction of credit records for college students, fully recording their various credit behaviors; Establish a nationwide unified personal credit reporting system to make it easier for college students with demand and repayment ability to obtain loans. For students with poor credit ratings, their loan eligibility will be revoked, further restraining the occurrence of dishonest behavior among college students.

Establish a multi departmental control mechanism, control chaos and prevent risks from the source, strengthen campus supervision, and create a good cultural atmosphere. Universities need to be vigilant, strengthen supervision of the surrounding environment, prohibit campus agents of various types of campus loans from entering the campus to conduct various financial lectures, distribute campus loan promotional brochures, and carry out various campus loan promotions in the name of activities. Students should be guided to distinguish right from wrong, make accurate judgments, make correct decisions, and stay away from various traps and deception. At the same time, a campus loan warning mechanism and disposal mechanism should be established to protect the safety of students' lives and property.

5.2. Establish a campus prevention and control system, strengthen security risk prevention

Universities have both the responsibility of maintaining campus security and stability, as well as the mission of cultivating students' growth and development. Each university needs to prioritize the risks and rectification of campus loans as a key task in maintaining campus stability at the current stage, actively improve management mechanisms, establish relevant control systems, and shoulder the dual responsibility of managing and educating students. We should mainly start from the following aspects:

1) Strengthen daily education for students, making prevention of campus loans one of the key work priorities, especially strengthening education on financial security for students. Strengthen campus network loan prevention and educational guidance. Efforts will be made to build a work system that is comprehensively coordinated by the Student Affairs Office and primarily responsible for each college; Efforts will be made to establish an information feedback team composed of student party branches, student unions, clubs, class groups, and other backbone members; Focus on building a reporting system for all teachers and students; Establish a timely mechanism for handling communication between parents and schools. By establishing a hierarchical system, the roles and responsibilities of various departments in campus loan risk prevention work are clearly defined, forming a collaborative force for campus loan risk prevention with each department taking responsibility.

2) Counselors and professional homeroom teachers should go deep into students, timely guide them to establish a reasonable consumption concept, and popularize safety knowledge. Through campus radio, WeChat official account, various WeChat groups of the class and other channels, they can understand the legal knowledge related to loans, and establish the awareness of credit loans and prevention awareness. Give full play to the expertise of teachers in economics, offer courses or lectures on financial management, actively convey real-time policies to students, popularize the latest relevant laws and policies, and help them scientifically distinguish financial risks.

3) The counselor collaborates with the Student Assistance Center to effectively carry out student assistance work. Counselors regularly visit student dormitories, promptly seek information from classmates, pay attention to students from economically disadvantaged families, carry out targeted poverty alleviation work, strengthen school funding and publicity, and help students find ways to solve economic difficulties and achieve self-reliance. The school should continuously improve various funding policies, pave the way and bridge with love, solve their worries, and enable disadvantaged students to study with peace of mind and make rapid progress.

5.3. Collaborative cooperation between universities and families, strengthening information communication and exchange

Strengthening communication and connection between schools and parents, exchanging student information between schools, and sharing student information, to achieve consistency in the mastery of student information between schools and families. This helps to comprehensively understand student consumption habits, and thus help cultivate good learning, consumption, and other behavioral habits in students in a targeted manner. At the same time, parents are urged to attach importance to and improve family education. Parents should help their children cultivate healthy consumption habits and financial awareness, guide them to learn independent thinking and rational consumption, establish correct values, and appropriately care about their children's financial situation to prevent potential problems.

5.4. Collaborative development of university platforms to promote business optimization and operation

Universities are the ones that have the best understanding of the student situation, apart from the student themselves. Students have a certain degree of understanding of their economic situation, consumption habits, self-discipline, development needs, etc. Campus loan platforms can understand the student's credit situation through universities at the beginning of borrowing, carefully review the student's loan conditions, and choose the best loan from the source. High quality borrowers are given lower loan interest rates, while low-quality borrowers are appropriately increased in interest rates, implementing differentiated pricing services. At the same time, they strictly comply with the national regulations on non bank institution loan interest rates, thereby ensuring the platform's commercial interests and providing preferential treatment for students.

5.5. Collaborative development between banks and enterprises, urging enterprises to strengthen self-discipline

The initial purpose of the "campus loan" business was to provide convenience for the student population and solve the problem of funding shortage for some students. The subsequent malignant incident of "campus loans" is rooted in the gap between the excessive consumption needs of contemporary college students and the lack of attention from formal financial institutions. Therefore, it is necessary to guide the development of legitimate campus loan platforms and not blindly curb and suppress them due to choking. Over time, this will not only fail to reduce the threat of non-performing platforms, but also increase the difficulty of supervision.

5.6. Establish a cooperation mechanism between banks and enterprises, and improve a sound online lending system

Cooperation between banks and enterprises is an inevitable trend in today's social development. Cooperation between banks and enterprises can not only expand the types of business of enterprises, but also ensure a more secure source of funds. Students will be more confident in obtaining loans from campus loan platforms, and companies will be more confident in transferring funds to students due to bank supervision. Collaboration between banks and enterprises can enable mutual benefit and win-win outcomes among banks, enterprises, and students. It can also reduce the probability of frequent malignant violent incidents in recent years, reduce the negative social impact caused by campus online loans, and be more conducive to the stability and development of the financial market.

The emergence of campus loans reflects the pursuit of capital's interests and the consumption needs of contemporary college students. The governance of campus loans is a process that requires the cooperation of multiple stakeholders. The governance of campus loans requires contemporary college students to enhance their awareness of safety and establish a rational consumption concept; Universities need to strengthen the cultivation of consumption concepts and financial management abilities among students, and establish and improve strict pre-education, early warning, and monitoring mechanisms. The governance of campus loans requires the government to strengthen the monitoring and management of campus loan platforms, create legal, compliant, safe, and fair lending platforms, improve legislation, and ensure that the management of campus loans has laws and regulations to follow; Financial institutions and platforms themselves should always prioritize integrity in their operations, consciously standardize business procedures, strictly review borrower information, and maximize risk avoidance for enterprises and college students. Through the joint efforts of multiple parties, it is believed that society can maximize the elimination of individual misfortunes and family tragedies caused by bad campus loans, promote the healthy and orderly development of campus loans, and build a good growth environment for the comprehensive development of college students.

6. Conclusion

Campus loans can bring convenience to college students, but they still carry certain risks. This article investigates, analyzes, and predicts the classification of campus loans from both soft and hard information perspectives. Firstly, starting from the aspect of soft information, the data from the campus loan questionnaire survey is statistically analyzed. Based on the statistical induction of data and prior knowledge, the information is deeply excavated, and reasonable reasons are analyzed to provide specific guidance for proposing and formulating campus loan risk control plans; Secondly, based on the obtained hard information, experiments were conducted using LGBM classification to demonstrate that using the application information of college students can predict the possibility of default in the future when applying for campus loans. This provides a feasible and accurate prediction method for predicting future campus loan data, and plays a real-time monitoring role in the risk of campus loans. Finally, based on both software and hardware information, and considering the actual situation of enterprises, governments, individuals, and relevant policies, corresponding prevention and control plans were proposed.

Acknowledgements

Social Science Planning Project of Shandong Province "Research on Campus Loan Risk Prevention Mechanism and Education Guidance in the Internet Context" (17CJYJ04)

References

[1] Wang, Sz., Li, Hz.: Research on the connotation and relationship between hard information and soft information. Journal of Information (04), 60-61+64 (2004).

[2] Angelini, E., Tollo, D G., Roli, A.: A neural network approach for credit risk evaluation. Quarterly Review of Economics and Finance 48(4), 733-755 (2008).

[3] Xu, Z., Zou, Cw.: Allocation of internal loan approval authority and incentive mechanism design in banks under the framework of hard and soft information: implications for SME financing issues. Financial Research (08), 1-15 (2010).

[4] Han, P., Xi, Ym.: Credit Risk Portfolio Forecasting Based on Fuzzy Neural Networks. Journal of Quantitative Economics and Technical Economics (05), 107-110 (2001).

[5] Zhang, J., Zhao, F.: Prediction of Default in Technology Credit for Small and Medium-sized Enterprises Based on Support Vector Machine. Statistics and Decision Making (20), 66-69 (2013).

[6] Wang, Xt., Zhao, Zx., Wang, Sht., Liu, Nn.: Research on default prediction of online loan borrowers based on machine learning. Service Science and Management 8 (1), 40-48 (2019).

[7] Zhou, Lf.: Research on Loan Default Prediction Based on Unbalanced Data Classification. Central South University, (2013).

[8] Luca Barbaglia, Sebastiano Manzan, Elisa Tosetti, Forecasting Loan Default in Europe with Machine Learning, Journal of Financial Econometrics, Volume 21, Issue 2, Spring 2023, Pages 569-596.

[9] Liang, L., Cai, X.: Forecasting Peer-to-Peer Platform Default Rate with LSTM Neural Network. Electronic Commerce Research and Applications 43(8), 100997 (2020).

[10] Arian, H., Seyfi, Sms., Sharifi, A.: Forecasting Probability of Default for Consumer Loan Management with Gaussian Mixture Models. arXiv preprint arXiv:2011.07906 (2020).

[11] Odegua, R.: Predicting Bank Loan Default with Extreme Gradient Boosting. arXiv preprint arXiv:2002.02011 (2020).

[12] Mi, C., Liu, J.: Role of text information in default evaluation of P2P lending platform—based on Multinomial Lasso-logistic Model. Mathematical Practice and Understanding (05), 78-91 (2019).

[13] Ma, Xj., Song, Yq., Chang, Bs.: Application research of P2P default prediction model based on CatBoost algorithm. Statistics and Information Forum (07), 9-17 (2020).

[14] Chen, Wh., Yu, Sm., He, J., Zhang, Zh., Li, Jc.: Reasons analysis and countermeasures of college students' campus loan based on two regression models. Journal of Normal Science 38(11), 34-38 (2018).

[15] Chen, Dy., Han, C.: A comparative study of online P2P lending in the USA and China. Journal of Internet Banking & Commerce 17(2), 1 (2012).

[16] Guo, Xw., Wang, X.: Hard information versus soft information: Which is more important? --Empirical study on loan default risk of non-listed smes. Quarterly Journal of Finance 009(002), 61-92 (2015).

[17] Jiang, Cq., Wang, Ry., Ding, Y.: P2P online lending default prediction method with soft information. China Management Science 025(011), 12-21 (2017).

[18] Zhang, Jp., Gu, X., Yang, C.: Analysis on risk factors of individual default in P2P campus loans. Finance and Accounting Monthly (06), 82-89 (2018).