Research on Rapid Selection of University Funding Objects Based on Social Big Data Analysis

Xiaoyan Xu(✉) and Yuliang Zhang

Data and Information Center, Wuxi Vocational Institute of Commerce, Wuxi 214000, China
15861562560@163.com

Abstract. With the sharp increase in the number of college students, the number of students who need financial aid also increases.How to quickly and accurately select university funding objects has become the key to achieve the goal of funding education. Therefore, this paper proposes a research on rapid selection methods of university funding objects based on social big data analysis. Based on the principles of systematicness, objectivity, scientificity and feasibility, we will build an index system for the selection of university funding objects, deeply mine the index data for the selection of university funding objects in the big data of social communications, build a pre-processing framework for the selection of index data, re sample the index data for the selection of university funding objects based on the SMOTE algorithm, and eliminate the adverse effects of unbalanced data. Set up a model for selecting university funding objects, formulate rules for selecting university funding objects, and realize rapid selection of university funding objects. The experimental results show that after the application of the proposed method, the corresponding maximum accuracy rate of the selection results of university funding objects is 98%, the maximum recall rate is 91%, and the maximum F value is 0.96, which fully confirms that the proposed method has better application performance.

Keywords: Big Data Analysis Means · Colleges and Universities · Quick Selection · Social Big Data · Funding Target.

1 Introduction

Since the expansion of college enrollment, more and more students have entered the college gate. On the basis of the steady growth of the number of students, the corresponding number of students in need of financial assistance is also increasing.Some new characteristics of poverty are shown in students from economically disadvantaged families. According to the data of the National Student Aid Management Center, in 2022, the government, universities and the society set up various university student aid policies to subsidize 4387 ordinary colleges and universities nationwide890000 person times, with a subsidy of 115.03 billion yuan. On the one hand, the university funding work has

L. Yun et al. (Eds.): ADHIP 2023, LNICST 547, pp. 436–451, 2024.
https://doi.org/10.1007/978-3-031-50543-0_30

been carried out like a raging fire, obtaining abundant resource support, helping students with financial difficulties from families to have a solid material foundation and a variety of policy help. All difficulties caused by family financial difficulties have been well curbed. Large scale, diversified and deep difficulties are gradually reducing, and university funding work has made some progress. The uneven distribution of students from economically disadvantaged families and the diversification of causes of poverty require that student funding needs to be more precise [1]. On the other hand, the economic and social development has promoted the continuous upgrading of the industrial structure, the substantial improvement of people's living standards, and the change of production and lifestyle, making the traditional material poverty alleviation need to combine the characteristics and requirements of the new era, on the material basis, strengthen the improvement of the spiritual world. University funding is a specific way of poverty alleviation for students. Since the 18th National Congress of the Communist Party of China, General Secretary, on the basis of socialism with Chinese characteristics, summarized the experience and lessons of anti-poverty, put forward new ideas for poverty alleviation in theory and practice, and created a new program for poverty alleviation with Chinese characteristics.

The main task of education reform and development in the "13th Five Year Plan" is to help more students with difficult families, complete the task of poverty alleviation, and achieve the goal of full coverage of college education, so as to achieve education fairness. Especially for some areas with relatively backward development, we should increase poverty alleviation capacity, improve poverty alleviation effect, and achieve targeted poverty alleviation. In the plan, it is emphasized that in order to ensure the full realization of education fairness, a university management information platform based on students' basic information should be established, and the information management system of civil affairs, schools, poverty alleviation and other departments should be organically combined, so as to open a "green channel" for poor students in rural remote areas, increase funding, and enable more students to enjoy the right to receive education and enjoy educational achievements. We will lift ourselves out of poverty. Under the background of big data, it is more and more convenient to obtain data that can reflect students' consumption characteristics and learning status. This research makes full use of the behavior data of students in school provided by various platforms under the background of big data, and introduces data mining methods into the funding work. On the one hand, it is conducive to the connotation, quality and method of financial aid,Improve the accuracy of the identification of funding objects. On the other hand, it uses the means of informatization to accurately identify funding objects, and fairly and reasonably allocate funding resources, so as to better solve the problem of malicious tampering with information in the current university funding work [2]. Accurate identification of funding objects is conducive to improving the effectiveness of funding work and better realizing the purpose of funding education.

Based on existing research results, the commonly used methods for rapid selection of university funding targets are the campus one card based method [3] and the impoverished student consumption information based method [4]. The former uses the campus all-in-one card data to study the laws of students' Consumer behaviour using the methods

of association analysis and clustering in data mining, so as to provide relevant suggestions for the distribution of school grants. The latter uses support vector machines and association rule methods in data mining to establish normal and abnormal consumption patterns for impoverished students based on data from the consumption information system of impoverished students in universities. Both of the above two methods have certain defects. The selection efficiency of university funding objects obtained is low, and the accuracy of selection results is poor, which affects the effect of funding education. Therefore, the research on rapid selection methods of university funding objects based on social big data analysis is proposed.

2 Research on Quick Selection Method of University Funding Objects

2.1 Construction of Index System for Selecting University Funding Objects

Under the background of big data, the construction of the index system for the selection of university funding objects should follow the following principles: systematic, objective, scientific and feasible, as shown in Table 1.

The indicator system for the selection of university funding objects is constructed according to the principles shown in Table 1, as shown in Table 2.

As shown in Table 2, the number of first level indicators, second level indicators and third level indicators is 3, 6 and 20, respectively, in the indicator system for the selection of university funding objects, which can comprehensively measure the actual situation of university funding objects and provide support for rapid and accurate selection of funding objects.

2.2 Preprocessing of Target Data for Funding Object Selection

Data mining is a technology that extracts useful information and patterns from a large amount of data, which can help understand data, discover hidden associations and trends, and provide support for subsequent analysis and decision-making. Data mining plays an important role in the preprocessing of funding target selection indicator data. It can also help clean data, select features, transform and integrate data, and discover association rules and patterns in the data. These steps help improve data quality, simplify analysis and decision-making processes, and provide more accurate and effective support for subsequent funding target selection.

Based on the above constructed index system for the selection of university funding objects, mine the index data for the selection of university funding objects in the big data of Social Communications, and preprocess them to provide assistance for subsequent research.

Data mining technology mainly includes five steps, as shown below:

(1) Data integration and cleaning: In the era of big data, data sources are diversified, and the data format is diverse and disorderly. Therefore, cleaning the acquired knowledge is of great significance for improving the accuracy of experiments.

Table 1. Principles for Selection of University Subsidies

Principle	Content description
systematicness	A complete index system for the selection of university funding objects should have sufficient coverage, including all aspects of students' information, from students' basic information, all-in-one card consumption, personal daily performance to academic achievements, so as to establish scientific management guidelines.The selected variables should be able to reflect the students' consumption level, consumption potential, poverty index, learning willingness, learning achievements, etc., without too many variables, leading to the prediction model being too complex.Therefore, no important indicator should be omitted in the selection of variables, and the situation should be reflected as completely, comprehensively and systematically as possible to avoid generalizing
objectivity	The indicators selected according to the research content can truly and objectively reflect the essential attributes of things.In the precise selection model of university funding objects, the important thing of the indicator system is to reflect the quality rather than the quantity.The selected variables should truly reflect the students' consumption level and learning attitude, so as to establish a working mechanism for selecting funding objects
Scientific	The design of each indicator system and the selection of evaluation indicators must be scientific. The indicators selected in this paper should objectively and truly reflect the characteristics of students' consumption behavior and learning attitude at school.The evaluation indicators should be representative, not overlapping, too complicated, but not too simple, to avoid information omission, information error and other phenomena
feasibility	The set indicator system is required to be highly operable, and the selected variables can truly reflect the essence of things, so the collected information should be easy to express and process, and the indicator setting should be concise, practical, and highly operable.The availability of data should be considered, so the selected variables must be easy to obtain to improve the accuracy and feasibility of the operation

(2) Data selection and conversion: The amount of data obtained after data cleaning is large, and not all data are available, so it is necessary to select research related data from them to reduce the complexity of the experiment. In addition, there may be inconsistent measurement standards between different features of the acquired data, such as discrete, continuous and other variables. Therefore, before using data mining algorithms for research, data variables need to be discretized and standardized.
(3) Data mining: mining different knowledge with different relationship types according to different algorithms. The common data mining algorithms include classification, clustering, feature extraction, association analysis, etc. According to the specific analysis of specific problems, select the appropriate algorithm.
(4) Mode evaluation: the validity of the data mining model is evaluated by such indicators as accuracy, accuracy, and sensitivity.

Table 2. Index System for Selection of University Subsidies

Primary indicators	Secondary indicators	Level III indicators
Individual consumption characteristics	Daily living expenses	Total supermarket consumption
		Total consumption of boiled water
		Total shower consumption
	Restaurant consumption	Number of restaurant consumption
		Total Meals
		Maximum single consumption amount
		Single minimum consumption amount
		Average meal amount
	Overall consumption	Total times of school consumption
		Total consumption
		Average consumption
		Remaining amount in the card
Learning attitude	academic record	Score ranking
	Daily performance	Book borrowing type
		Total books borrowed
		Times of entering and leaving the library
		Total amount of dormitory
		Total amount out of dormitory
		Proportion of books borrowed from professional courses
Funding	a grant	Access to financial aid

(5) Knowledge representation: use visual knowledge to show users the results of data mining.

The process of in-depth data mining for selecting indicators of university funding objects is shown in Fig. 1.

The process shown in Fig. 1 is used to mine the index data of university funding object selection. Due to the influence of various factors, there are redundancy, missing and other phenomena in the mining data. If it is directly applied, it will affect the accuracy of the selection of university funding objects. In addition, there are many student consumption scenarios, and consumption data is accumulating at an alarming rate every day. There are sensitive information leakage, redundant and useless data, missing key data, abnormal data and other problems in the data, so it is necessary to establish a unified and deployable

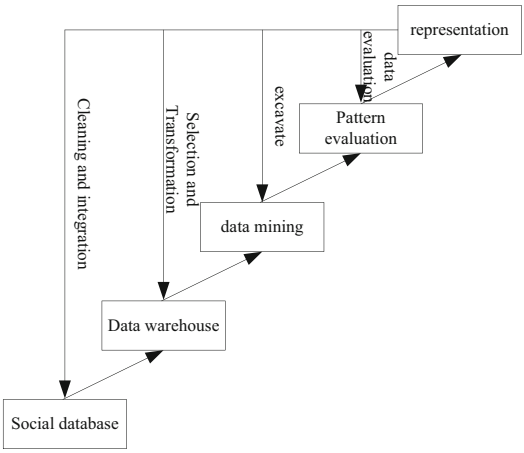

Fig. 1. Schematic diagram of data mining process for selecting indicators of university funding objects

pre-processing framework to apply to the large amount of data. It is necessary and practical to have many data dimensions. The data preprocessing framework in this paper mainly includes four modules: data desensitization, data cleaning, data specification and data integration.The data preprocessing framework is shown in Fig. 2.

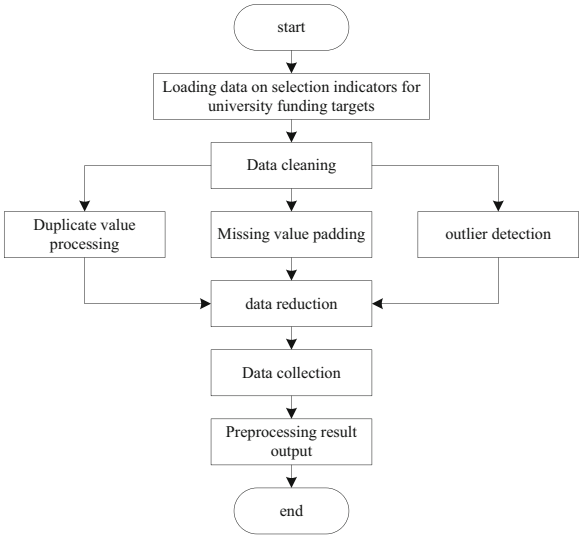

Fig. 2. Framework of data pre-processing for selecting indicators of funding objects

Through investigation and analysis of data, it can be seen that outliers are the abnormal data that has the greatest impact on the selection of university funding targets, so it is necessary to accurately detect and process them. The purpose of calculating the

distance between the selection indicators of funding targets is to evaluate the similarity or difference between different targets. By calculating the distance, the degree of difference between different objects in indicator data can be quantified, thereby achieving outlier detection [5]. The formula for calculating the distance between indicator data is

$$d_{ij} = \alpha^o * \sqrt{\sum_{i,j=1}^{n} \left| x_i - x_j \right|^{\alpha^o}} \tag{1}$$

In formula (1), d_{ij} It represents the index data of the selection of university funding objects x_i And x_j Minkowski distance. Minkowski distance as a method to calculate the distance between selected index data of funding objects, has the advantages of flexibility, adjustable robustness, interpretability and Foundations of mathematics. These advantages make Minkowski distance a commonly used and reliable distance measurement method, which has wide applications in funding object selection and other data analysis tasks; α^o It represents a variable parameter. The value range is [0, 1]; n It represents the total number of index data for the selection of university funding objects.

According to formula (1), measure the average Minkowski distance between the indicator data x_i for selecting university funding targets and all other data. The calculation formula is

$$\overline{d}_i = \frac{\sum_{j=1}^{n-1} d_{ij}}{n-1} \tag{2}$$

In formula (2), $\overline{d}_i$ It represents the index data of the selection of university funding objects x_i Average distance from all other data.

Calculate the result with formula (2)$\overline{d}_i$ To determine the index data for the selection of university funding objects x_i Whether it is an outlier, the specific determination rules are as follows:

$$\begin{cases} \overline{d}_i \le \beta^* & \text{Normal Value} \\ \overline{d}_i > \beta^* & \text{Outlier} \end{cases} \tag{3}$$

In Eq. (3), β^* It represents the outlier determination threshold, which needs to be set according to the actual selection of indicator data [6].

The outlier detection results of university funding object selection index data obtained according to the above rules are shown in Fig. 3.

As shown in Fig. 3, the data points inside the green dotted circle are outlier detection results, which fully indicates that the outlier detection effect of the selected indicator data of university funding objects is better. Delete them to reduce the proportion of abnormal data in the selected indicator data.

On the basis of the above outlier deletion processing of university funding object selection index data set, each index data is accurately clustered using hierarchical clustering algorithm to provide a basis for subsequent unbalanced data processing.

Hierarchical clustering is mainly divided into top-down hierarchical splitting method and bottom-up hierarchical agglomeration method, which are described in turn below.

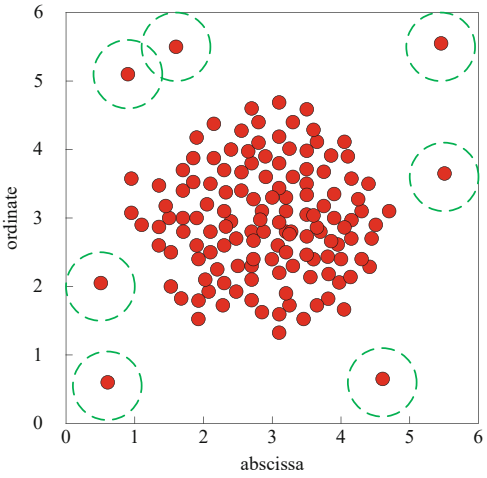

Fig. 3. Example of outlier detection results of university funding object selection index data

The representative of hierarchical agglomeration is AGNES algorithm. For sample size of n Datasets for $\{x_1, x_2, \cdots, x_n\}$ Given the definition of distance and connection mode, the specific steps of the algorithm are as follows:

Initial step: each point is a class, and the dataset is divided into n Classes, namely $C_i \in \{x_i\}$;

Step 2: Calculate the distance matrix between different classes D;

Step 3: find the two nearest classes and merge them into one class.

Step 4: Repeat steps 2 and 3 until all data belong to a class or meet a termination condition[7].

The complexity of AGNES calculation is $O\left(n^2 \log_2 n\right)$. (n2logn) is not suitable for large-scale data sets, that is, not suitable for this study. The DIANA algorithm, which is the representative of hierarchical splitting, is completely different from AGNES algorithm in this process. Gradually descending layering[8]is adopted at one point.

First, all data are classified into one category, that is $A_1 = \{x_1, x_2, \cdots, x_n\}$;

Secondly: it is necessary to measure the length between all points, and select the average maximum length leading to different points (mark S), take this point as a new starting point, namely $S \in A_2$;

Then: face every point $x_i \in A_2$, the distance calculation formula is defined as

$$D_i = d_{avg}(x_i, A_1) - d_{avg}(x_i, A_2) \tag{4}$$

In Eq. (4),D_i Represents a class A_1 And A_2 Distance;$d_{avg}(x_i, A_1)$ It represents the selection of indicator data x_i And class A_1 Distance between;$d_{avg}(x_i, A_2)$ It represents the selection of indicator data x_i And class A_2 Distance between.

According to the calculation result of formula (4), select the largest D_i. If D_i If it is greater than 0, then $x_i \in A_2$;

Finally: cycle the previous step until all D_i If it is less than 0, it is regarded as turning a group into two parts.

After all the processes, the DIANA algorithm converts a sample data into X The dataset of is divided into two groups A_1 and A_2, repeat the above links indefinitely, split into different forms, until all the data together or all of them reach the critical value, clustering work is suspended.

Record the index data of college funding object selection after the above processing as $Y = \{Y_1, Y_2, \cdots, Y_n\}$, where,$Y_i$ It refers to the No i The data set of selected indicators provides assistance for subsequent research.

2.3 Unbalanced Data Processing Based on SMOTE

Pre processing results of index data selected by the above university funding objects $Y = \{Y_1, Y_2, \cdots, Y_n\}$ As a basis, through in-depth exploration and analysis, we can see that the data used in the selection of university funding objects is naturally unbalanced, which will have a great adverse impact on the selection results of university funding objects, so it is necessary to deal with it.

SMOTE algorithm is an intelligent oversampling technology proposed by Chawla et al. This method is not to increase the sample size by mechanical copying, but to synthesize new positive samples through certain rules to balance the data. It is a heuristic sampling algorithm. The main idea of this algorithm is to use linear interpolation and K Nearest neighbor method, where the distance from the sample is small, uses linear interpolation to synthesize minority samples, which effectively alleviates the problem of data over fitting [9]. The main steps of SMOTE algorithm are as follows:

Step 1: Select indicator data samples for each Y_i, find it first K Neighbor samples of the same type and the smallest distance from them, so that the upward sampling rate is M, from here K Randomly select samples from neighbor samples with small distance m, and recorded as $\{y_1, y_2, \cdots, y_m\}$;

Step 2: In minority data samples Y_i And $\{y_1, y_2, \cdots, y_m\}$ The random linear interpolation method is used to synthesize new minority data samples, and the expression is

$$Y_{new} = Y_i + rand * (y_i - Y_i) \tag{5}$$

In formula (5),*rand* It represents any random value between 0 and 1.

Step 3: Insert the new composite sample into the dataset to form a new dataset.

The sampling effect of the index data set of university funding object selection based on SMOTE algorithm is shown in Fig. 4.

With the new dataset above $Y' = \left\{Y_1', Y_2', \cdots, Y_n'\right\}$ To determine the resampling frequency of each data set, the formula is

$$f_i = \frac{Y_i' - Y_i}{\chi^{\Theta}} \tag{6}$$

In formula (6),f_i It means to select indicator data set Y_i Resampling frequency of;χ^{Θ} It represents the auxiliary parameter of resampling frequency calculation, and the value range is [0, 1].

The SMOTE algorithm shown above is used to resample the index data set of university funding object selection, eliminate the adverse effects of unbalanced data, and

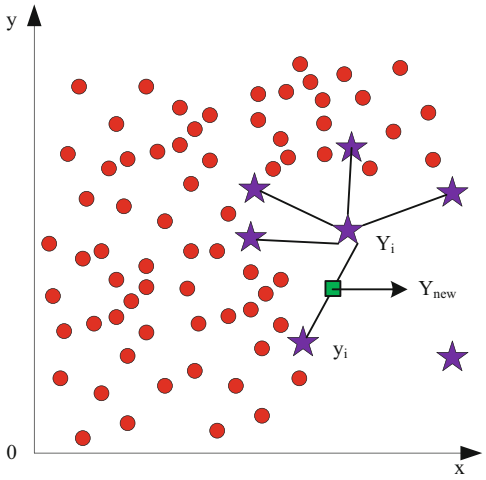

Fig. 4. Sampling Effect Diagram of Index Data Set for Selection of University Funding Objects

make sufficient preparations for obtaining the final results of university funding object selection.

2.4 Obtaining the Results of Selecting University Funding Objects

Record the index data of university funding object selection after unbalanced data processing as $Z = \{Z_1, Z_2, \cdots, Z_n\}$ In combination with the index system of university funding object selection built in Sect. 2.1, a model of university funding object selection is built, with the expression as

$$\psi_i = \frac{7 \times \delta^2 \times \sqrt{Z_i^2}}{\sum_{i=1}^{n} Z_i \times R_i + \varepsilon_i} \tag{7}$$

In Eq. (7),ψ_i It represents the value of the objective function for the selection of university funding objects;R_i It refers to the selection index of university funding objects;ε_i It indicates the adjustment parameters of the selection error of university funding objects;δ^2 It represents the standard factor for the selection of funding objects.

Calculate the result with formula (7)ψ_i Based on, the selection rules of university funding objects are formulated as follows:

$$\begin{cases} \psi_i \geq \zeta^{\tau} & \text{Funding Object} \\ \psi_i < \zeta^{\tau} & \text{Unfunded Object} \end{cases} \tag{8}$$

In Eq. (8),ζ^{τ} It refers to the decision threshold [10] for the selection of university funding objects.

Through the above process, we obtained the results of the selection of university funding objects, realized the rapid and accurate selection of university funding objects, and provided assistance for the realization of the purpose of funding education.

3 Experiment and Result Analysis

3.1 Select the Indicator Data Set to Determine the Resampling Frequency

The data sources of university funding targets are as follows:

(1) University's own data: Comprehensive universities usually have their own student funding offices or relevant departments responsible for collecting and managing student funding information. These departments may maintain students' personal information, family background, economic status, and other information to determine whether they meet the funding requirements.
(2) Government agencies: Government education departments or related institutions may collect and manage data related to student funding in universities. These institutions may collaborate with comprehensive universities to share student funding information in order to better manage and allocate funding resources.
(3) Survey and research data: Independent survey and research institutions or academic research teams may conduct surveys on university student funding to collect relevant data. These data may include information such as students' economic status, funding needs, and types of funding.
(4) Cooperation with other universities: In addition to comprehensive universities, other universities can also be sources of data on university funding targets. Cooperation and information sharing can be carried out between different universities to better understand and evaluate students' funding needs.

In order to avoid the adverse effects of unbalanced data, a method based on SMOTE algorithm is proposed to resample the selected index data set. Determine and select the resampling frequency of the indicator data set according to Formula (6) to facilitate the subsequent experiments, as shown in Table 3.

As shown in the data in Table 3, the determined resampling frequency is used to collect the required experimental data in the index data set of university funding object selection, so as to maximize the application performance of the proposed method.

3.2 Selection of Evaluation Indicators

In general, there are many situations in the selection of university funding objects, as shown in Table 4.

As shown in Table 4, TP refers to the number of samples originally funded and selected as funded objects;FN represents the number of samples originally funded and selected as non funded objects; FP refers to the number of samples that were originally non funded and selected as funded objects; TN refers to the number of samples that were originally non funded objects but were selected as non funded objects.

According to the needs of university funding object selection, the accuracy rate, recall rate and F value are selected as evaluation indicators, and the expression is

$$\begin{cases} Q_1 = \frac{TP}{TP+FP} \\ Q_2 = \frac{TP}{TP+FN} \\ F = \frac{(1+\mu_1^2)Q_1 \times Q_2}{\mu_1^2 Q_2 + Q_1} \end{cases} \tag{9}$$

Table 3. Resampling Frequency of Selected Indicator Data Set

Secondary indicators	Level III indicators	Resampling frequency (Hz)
Daily living expenses	Total supermarket consumption	44
	Total consumption of boiled water	30
	Total shower consumption	50
Restaurant consumption	Number of restaurant consumption	59
	Total Meals	65
	Maximum single consumption amount	42
	Single minimum consumption amount	23
	Average meal amount	20
Overall consumption	Total times of school consumption	47
	Total consumption	58
	Average consumption	59
	Remaining amount in the card	55
academic record	Score ranking	70
Daily performance	Book borrowing type	84
	Total books borrowed	51
	Times of entering and leaving the library	56
	Total amount of dormitory	40
	Total amount out of dormitory	38
	Proportion of books borrowed from professional courses	69
a grant	Access to financial aid	52

Table 4. Selection Results of Funding Objects

	Funding objects	Non funded objects	total
Funding objects	TP	FN	P
Non funded objects	FP	TN	N
total	P'	N'	P + N

In Eq. (9),Q_1 It indicates the accuracy of the selection of university funding objects;Q_2 It refers to the recall rate of university funding object selection;F It represents the F value corresponding to the selection result of university funding objects.F It is a comprehensive

evaluation standard,μ_1 In order to adjust the coefficient between recall rate and accuracy rate, it is usually taken as 1.If we can see from the expression,F The criteria can correctly measure the performance of the proposed method, if the recall rate Q_2 And accuracy Q_1 High, description F The value is high, which can better identify minority features. When colleges and universities accurately identify funding objects,F A high value indicates that the precise funding prediction model established can identify more funded students and more accurately identify funding objects.

3.3 Analysis of Experimental Results

According to the determination results of resampling frequency of the above selection index data set and the selection results of evaluation indicators, the quick selection method of university funding objects based on campus all-in-one card and the quick selection method of university funding objects based on poor students' consumption information are set as comparison methods 1 and 2 to carry out a comparative experiment of university funding object selection. The specific analysis process of the experimental results is as follows:

3.3.1 Analysis on the Accuracy of the Selection of University Funding Objects

Table 5 shows the accuracy rate of selecting university funding objects through experiments.

Table 5. Accuracy of Selection of University Subsidies (Q_1/%)

Experimental group	Propose method	Comparison method 1	Comparison method 2
1	89	56	45
2	94	45	56
3	95	52	43
4	98	41	59
5	84	52	53
6	85	58	44
7	78	47	47
8	85	59	59
9	81	41	52
10	76	55	50

As shown in the data in Table 1, after the application of the proposed method, the accuracy rate of university funding object selection is 76% ~ 98%; After the application of comparison method 1, the accuracy rate of the selection of university grants is 41% ~ 59%; After the application of comparison method 2, the accuracy rate of university funding object selection is 43% ~ 59%. Through comparison, it is found that after

the application of the proposed method, the accuracy rate of university funding object selection is far higher than that of comparison method 1 and comparison method 2, and the maximum accuracy rate in the fourth experimental group is 98%.

3.3.2 Analysis on Recall Rate of University Funding Object Selection

See Table 6 for the recall rate of university funding objects selected through experiments.

Table 6. Recall rate of university funding object selection (Q_2/%)

Experimental group	Propose method	Comparison method 1	Comparison method 2
1	78	55	36
2	68	45	39
3	89	40	40
4	88	36	41
5	74	45	31
6	71	42	29
7	76	41	45
8	89	48	48
9	90	50	47
10	91	35	59

As shown in the data in Table 6, after the application of the proposed method, the selected recall rate of university funding recipients ranges from 68% to 91%; After the application of comparison method 1, the selected recall rate of university grants is 35% ~ 55%; After the application of comparison method 2, the selected recall rate of university funding objects is 29% ~ 59%. Through comparison, it is found that after the application of the proposed method, the selected recall rate of university funding objects is much higher than that of comparison method 1 and comparison method 2, and the maximum recall rate of the 10th experimental group is 91%.

3.3.3 F-value Analysis of the Selection Results of University Funding Objects

According to Q_1-Q_2 The curve shows that researchers often hope for accuracy Q_1 and recall rate Q_2 Both of them are very high, but in fact they are contradictory. The above two indicators are contradictory and cannot achieve double high. For example, if you want a high accuracy rate, you must sacrifice some recall rates. To get a high recall rate, we must sacrifice some accuracy. But usually, a new indicator can be defined according to the balance point between them F Value, which can consider the accuracy rate and recall rate at the same time, so that both can reach the highest and achieve a balance.

The F value of university funding object selection results obtained through experiments is shown in Fig. 5.

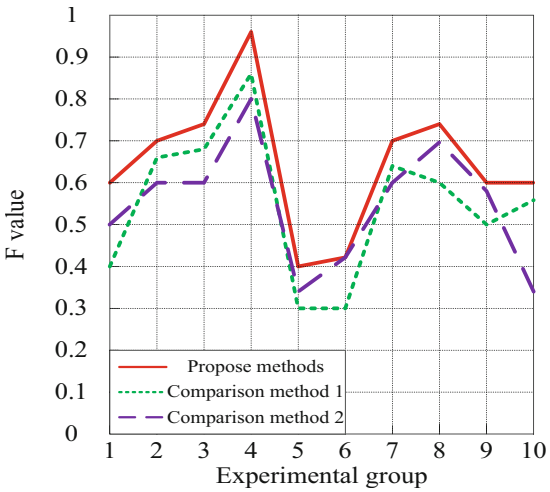

Fig. 5. Schematic diagram of F value of university funding object selection results

As shown in the data in Fig. 5, after the application of the proposed method, the F value range of the selected university funding objects is 0.4 ~ 0.96; After the application of comparison method 1, the F value range of university funding object selection results obtained is 0.3 ~ 0.86; After the application of comparison method 2, the F value range of the selected university funding objects is 0.34 ~ 0.8. Through comparison, it is found that after the application of the proposed method, the F value of the selection result of university funding objects is much higher than that of comparison method 1 and comparison method 2, and the maximum F value of 0.96 is obtained in the fourth experimental group.

4 Conclusion

In the context of big data, introducing big data into the work of funding education, and using data mining methods to analyze students' campus behavior data can objectively provide reliable credit guarantee, reduce human intervention, maximize the accuracy of data, and provide a new method for accurately identifying funding objects. This paper mainly studies the relevant issues when using data mining technology to accurately identify funding objects, and conducts in-depth discussions around the improvement of the target system for accurate identification of funding objects and the imbalance of data when selecting identification methods. First, in view of the status quo of precise identification indicators of university funding objects established in the context of big data, it is proposed that students' daily performance information should be integrated and their learning attitude should be comprehensively considered. Therefore, daily performance indicators are introduced into learning attitude, and a precise identification indicator system of funding objects in line with the background of big data has been established. Secondly, when establishing the prediction model for precise identification of university funding objects, it was found that the unbalanced data made it difficult to accurately

identify the behavior characteristics of the funded students. Based on this, this paper proposed to use the SMOTE resampling algorithm to balance the data set in the data pre-processing stage, which enhanced the prediction accuracy of classification, so as to more accurately identify the funding objects. The experimental data shows that the proposed method effectively improves the accuracy, recall rate and F value of university funding object selection, and can provide more effective method support for funding object selection.

Aknowledgement. 2022 University Philosophy and Social Sciences Research Project: Research on the Realistic Dilemma and Technical Demands of Precise Funding in Universities from the Perspective of Big Data (2022SJYB1050).

References

1. Behera, R.K., Jena, M., Rath, S.K., et al.: Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. Inf. Process. Manage. **58**(1), 102435 (2021)
2. Weimar, D., Soebbing, B.P., Wicker, P.: Dealing with statistical significance in big data: the social media value of game outcomes in professional football. J. Sport Manag. **35**(3), 266–277 (2021)
3. Zhang, Z.: Research on the identification of poor college students based on campus card data. Modern Economic Information **12**(13), 12–19 (2017)
4. Cao, L.Z.: Analysis and research on identification of poor students in art vocational colleges based on campus all-in-one card consumption big data. Comput. Knowl. Technol. **16**(30), 3–10 (2020)
5. Guo, W., Gao, J., Tian, Y., et al.: SAFS: object tracking algorithm based on self-adaptive feature selection. Sensors **21**(12), 4030 (2021)
6. Radajewski, D., Hunter, L.: An innovative data processing method for studying nanoparticle formation in droplet microfluidics using X-rays scattering. Lab Chip **21**(22), 4498–4506 (2021). https://doi.org/10.1039/D1LC00545F
7. Juneja, P., Garg, R., Kumar, P.: Uncertain data processing of PMU modules using fuzzy petri net. J. Intell. Fuzzy Syst. **41**(3), 1–13 (2021)
8. Kamiński, M., Szabat, K.: Adaptive control structure with neural data processing applied for electrical drive with elastic shaft. Energies **14**(12), 3389 (2021). https://doi.org/10.3390/en14123389
9. Lee, W.J., Staneva, V., Mayorga, E., et al.: Echopype: enhancing the interoperability and scalability of ocean sonar data processing. J. Acoust. Soc. Am. **149**(4), A63–A63 (2021)
10. Yun, S., We, G.: A python package for intennediate data processing in UAV based plant phenotyping. Remote Sens. **13**(13), 2622 (2021)