Application of Data Mining Technology Based on Weka in Student Management

Yang Li(✉)

Wenhua College, Wuhan 430074, Hubei, China
liyang3713@dingtalk.com

Abstract. This paper studies the application of data mining technology based on Weka in student management. As an open data mining platform, Weka collects a large number of machine learning algorithms that can undertake the task of data mining, including data preprocessing, classification, regression, clustering, association rules and visualization on the new interactive interface. Student work is the central work of the school, and student management is the top priority of school management. Reflecting on the management of school students, I personally believe that the following aspects still need to be further strengthened: conduct in-depth research on the decision tree analysis method in data mining technology, deeply analyze the environment and resources of the school, correctly evaluate their own level, make their own value orientation, and put C4 The algorithm is applied in student management, constructs the student psychological state model based on student personal information, and obtains some relevant laws. Make the educational behavior complete the rational leap, form the correct teaching management thought, and lay the foundation for quality education to a higher level. Practice has proved that this method improves the efficiency and quality of students' work, and provides a scientific reference for students' management and guidance system.

Keywords: Weka software · Data mining technology · Student management

1 Introduction

Data mining technology through in-depth analysis of information to find out the potential relationship and law between data. Data mining technology is applied to student management. Through the analysis of student data, the hidden internal relationship between students' school performance and various factors is discovered, so as to predict students' behavior and guide students' management.

Quality education is the main melody of current education. School is a place to educate people and a place to cultivate people. To make quality education effective, we must also carry out effective student management. Only effective student management can ensure the smooth progress of quality education. Advanced teaching and student management is to meet the requirements of society and the need of education and training multi-standard talents.

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 225–235, 2023.
https://doi.org/10.1007/978-3-031-23950-2_25

Society is a colorful group. It needs talents of various types, specifications, levels and characteristics. The arrival of knowledge economy has put forward new requirements for education. It requires that education is not only a comprehensive education, but also to realize the unity of scientific education and humanistic education, general education and training, and realize the coordination of knowledge, ability, morality and emotion. While paying attention to cultivating the creativity of the educatee, it also needs to cultivate his coordination ability and cooperation consciousness with the surrounding world; Moreover, education is required to be creative, diverse, open and personalized. As a student manager of a school, we must stand at the commanding height of the times, focus on the 21st century, deeply analyze the environment and resources of the school, correctly evaluate our own level, and make our own value orientation, so as to outline a viable development strategy and basic goal in line with our own reality, so as to make a rational leap in educational behavior, Form a correct teaching and student management thought and lay a foundation for quality education to a higher level. Rationality is the crystallization of wisdom and the embodiment of rationality. The management of teaching students must pay attention to rational innovation. Student management should be good at active learning and lead a group of teachers who are good at learning. Student management should master advanced ideas in continuous learning experience, change educational ideas in learning, establish a correct outlook on education by learning from the experience of various schools at all levels, and obtain a fresh outlook on Teaching in the training, teaching and research and teaching activities with teachers.

Only by adhering to long-term learning can student management skillfully combine practical experience with theoretical wisdom, and create new methods, new languages, new ideas and new experience with school characteristics. Teaching and student management should be emotional. Convince people with reason and move people with emotion [1]. The ingenious combination of reason and emotion can achieve good results. Pay attention to the close relationship between schools and society. Under the background of today's market economy, student management needs innovation and communication with people from all walks of life. In the process of communication, we should adhere to the two principles of being conducive to the development of the school and abiding by laws and regulations. Correctly handle the relationship with the superior leaders, correct the relationship between the superior and the subordinate, learn the way of transposition thinking, learn to be good at expressing opinions, and obtain the help of the superior leaders and the support of the subordinate colleagues. Make student management get twice the result with half the effort.

School based student management is the trend of contemporary school student management reform. Due to the long-term constraints of the "external control" school student management system, student schools generally lack the tradition and practical experience of school-based student management, and the institutionalized operation of school-based student management still needs a gradual realization process.

With the advent of knowledge economy and information age, social competition is becoming increasingly fierce, and parents pay more and more attention to their children's education. They not only require schools to improve the quality of education, but also have a growing desire and voice to participate in school student management. With the development of community education, school student managers gradually realize that school and community are closely related. The survival and development of schools must rely on, base on and serve the community. All these not only exercise the school-based student management ability of school student managers, but also enhance the subject consciousness of all sectors of society to participate in school-based student management [2]. The innovative concept of teaching and student management will bring new vitality to the cause of education. With the continuous reform of the teaching system in Colleges and universities in recent years, especially the development and deepening of the credit system and course selection system, the daily management of students is becoming more and more heavy and complex. In the daily management of students, there is a lot of work to count the information of students in all aspects, such as student status, achievement, award evaluation, dormitory, teaching and employment, etc. Although the functions of querying and processing the information of students' growth can not be realized efficiently according to the existing information of people and rules, it is impossible to find and process the information of students' growth.

Student education is the key basic education stage of life. Education and teaching management is an important part of student management, which marks the comprehensive strength and school running level of the school. Efficient education and teaching management is a powerful guarantee for students' good teaching quality. Teaching management refers to the management of both teachers and students according to the requirements of social development and the development characteristics of students [3]. The purpose of teaching management is to coordinate the relationship between teachers and students in teaching practice and achieve good teaching effect. At present, the rapid development of economy and culture has promoted the development of all walks of life. With the proposal of comprehensive quality education and the implementation of the new curriculum reform, teaching ideas and teaching methods have changed. It is of great practical significance to explore the effective strategy of student education and teaching management.

2 Related Work

2.1 Weka Software

Weka is one of the platforms of data mining. Weka integrates many algorithms, which are mainly used in data mining, such as face recognition, text recognition, genetic engineering and so on. Different algorithms have different effects [4]. The main developer of Weka is from Waikato University in New Zealand. It is a free, non-commercial, open source machine learning and data mining software based on Java environment. Figure 1 below shows the Weka software interface.

Weka can preprocess, classify, cluster, association rules, attribute selection and visualization of data. Weka's source code is written in Java. All its jar packages can also be imported into Java projects and directly call APIs. Weka can be opened Xls file (Excel file) can be connected to the database (MySQL database is connected here). Weka data format: ARFF (attribute relation file format) file is shown in Fig. 2 below.

There are four commonly used test options: used as training set, used as test set, cross test (10 times cross test by default), and proportional segmentation (dividing the data

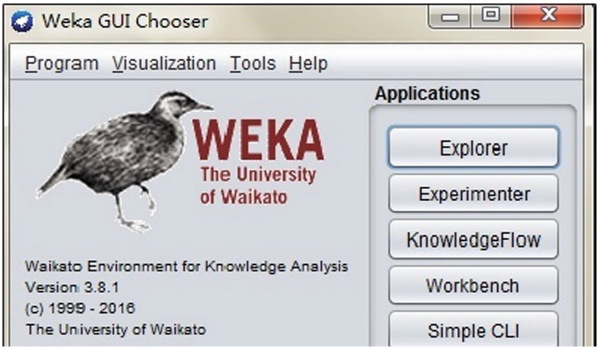

Fig. 1. Weka software interface

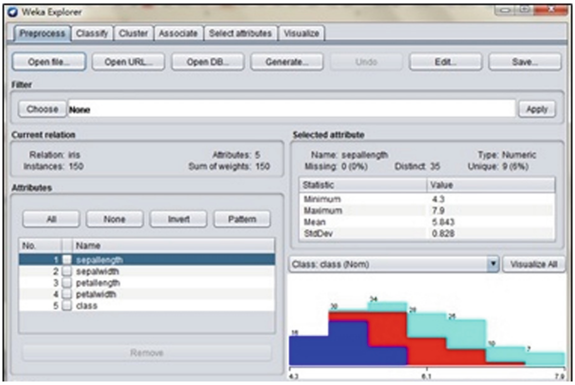

Fig. 2. Weka data format

set into training set and test set according to a certain proportion). In addition, there are some options for classifier output, such as evaluation index, choosing to output classifier code in Java language, etc. [5].

All the current classification results are displayed in the result list. Right click a result in the list and some options will appear, such as visualizetree (graphical decision tree), visualize classifier errors (the classification results are represented in a scatter diagram, the correct classification instances are represented as small crosses, and the wrong classification instances are represented as small hollow squares), as well as some other options. Open the explorer interface, click the preprocessing tab, import the nominal weather data set (weather. ARFF), enter the classification tab, select j48 classifier, test options, select as the training set, the decision tree will be generated, and the relevant output will be output in the classifieroutput. Select test options again, select as the test set strategy, then find output predictions in more options, and select plain text to output the test situation.

2.2 Data Mining Principle

Data is the most valuable commercial capital and the prerequisite of data mining. Valuable information sources can be found in enterprises and have a wide range of boundaries. For example: internal and external data sources, both structured and unstructured.

Internal data sources (such as databases, reports and spreadsheets) store a large amount of data. External databases and various Internet resources are also excellent ways to obtain business value. They can be combined with internal resources to better insight into the problems being studied.

The quality of data is the key to effective data mining. Without proper data, even if the best data mining experts are employed, it is impossible to obtain any valuable knowledge. The most important data quality requirements are:

Relevance - whether the data is applicable to the current problem;
Validity - whether the data follows the current environmental laws;
Integrity - noisy data or missing data are two major factors affecting data integrity;
Consistency - Information Island is the main reason for data inconsistency. This problem can be solved by using integrated information system · display visualization; Portals, aggregation applications, dashboards, charts, spreadsheets, etc.;
Trust - use reliable data sources to effectively process subjective data;
Timeliness - minimizes the time delay between data collection and data processing.

In short, data mining is to extract or "mine" knowledge from a large amount of data. Data mining should be more correctly named "mining knowledge from data". Excavation is a very vivid term, which captures the characteristics of the process of finding a small amount of gold bullion from a large number of unprocessed materials [6].

Data mining is the application of a series of technologies to extract information and knowledge of interest from large databases or data warehouses. These knowledge or information are implicit, unknown in advance and potentially useful. The extracted knowledge is expressed in the form of concepts, rules, laws, patterns and so on. It can also be said that data mining is a kind of deep-seated data analysis.

Data mining, also known as data mining and data mining, is to extract rules (or patterns) that can be interpreted as knowledge from massive data according to the established business objectives, including association rules, feature rules, distinction rules, classification rules, summary rules, deviation rules, clustering rules, etc. Most of them are based on data mining, test and learning. The methods corresponding to these technologies are often confusing to both novices and experienced data analysts [7].

It can be considered that the data mining method mainly consists of three parts: model representation, model evaluation and search. Model representation is a language used to describe patterns that can be found. If this representation is too limited, no amount of time and examples can produce an accurate model for the data. It is important for a data analyst to fully grasp the representative assumptions that may be implied in a particular method. Similarly, it is also very important for an algorithm designer to clearly express what representative assumptions a specific algorithm can make. Note that with the enhancement of the model representation ability, the risk of over fitting of the model to the training data also increases, resulting in the reduction of its prediction accuracy of unknown data.

2.3 Student Management

With the continuous promotion of the new curriculum reform, we now attach great importance to the subject status of students, and the content of student management is the main student management resource, which has a decisive impact on students' education and student management. Therefore, teachers must reasonably arrange the content of student management. In the process of student management, according to students' learning foundation and students' cognitive ability, Arrange the corresponding learning plan, and also consider the acceptance of students, so that teachers can present some student management contents in a more specific way in the process of student management, such as using multimedia to present some knowledge in the form of pictures, pictures and videos, which can deepen students' learning impression and enhance students' school efficiency [8], In addition, the self-management ability of students in this stage is relatively poor. In the process of explaining the student management content, teachers must make some student management content live, so as to stimulate students' learning initiative and enthusiasm, so as to improve students' learning efficiency and continuously enhance students' abilities in all aspects.

The improvement of students' management quality is not achieved overnight, which requires the cooperation and common development of all forces. According to the experience of the effectiveness of student management in education summarized previously, to improve the quality of student management, we must require the whole society, the whole school, teachers, parents and students to actively cooperate in the process of daily student management, give full play to their strengths in improving the quality of student management and complement each other's weaknesses [9], And then give full play to the advantages of each party. The responsibility of head teachers is more arduous. As a bridge connecting schools and family students, we should give full play to our responsibilities and urge all parties to jointly improve the quality of student management. In daily learning, it is needless to say that it is important to urge students to be individual. In the educational work, we should also timely feed back the problems encountered in the process of educational student management to the school superiors, so as to make a specific analysis of specific problems, and it is best to put forward suggestions or solutions. After all, the head teacher is the direct receiver of the problems, and can better understand the seriousness of the problems, so as to lay the foundation for the school to feed back the problems in time. For students' family learning environment, we should also be diligent in communication and work with parents to help their children grow and develop.

3 Application of Data Mining Technology Based on Weka in Student Management

Most college student management workers show a certain weak awareness of informatization in the specific work process, that is, they do not understand the development trend of informatization and the relationship between informatization and student management. Affected by many factors, some colleges and universities began to practice student management informatization, but the practice only stayed in the primary stage, did not make in-depth attempts and exploration, and the informatization effect was weak.

Student management is a relatively complex work, which is long-term and systematic. The practice process is accompanied by a large amount of data. Through observation and analysis, these data have the characteristics of low degree of structure. They are mainly composed of unstructured data and semi-structured data. Unstructured data mainly includes text, video, documents and pictures, Semi structured data mainly involves e-mail, resource database and reports. However, there are still some deficiencies in mining and processing, and no in-depth analysis and exploration has been made, which has affected the degree of data structure to a great extent. A member variable uses an array of member variables to save the corresponding data [10]. All contents depend on the operation of the class object, and the data saved in each management class object is different.

The difference at present is for better reusability. Like the later data, it is for the unity of data. The powerful function of Weka software provides a platform for us to deeply study student information. Based on the existing student information data, we preprocess the data, then analyze and make decisions on the processed data through Weka software, and finally display the data and laws in the form of tables and graphics, and finally build a perfect student management and guidance system (see Fig. 3).

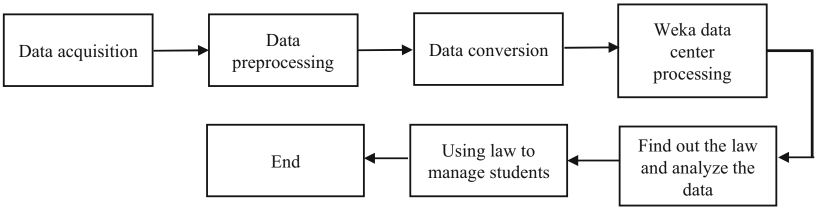

Fig. 3. Student guidance management system

First, use big data thinking in the student course selection system. Due to the relatively large distribution of majors in Colleges and universities, there are more professional courses and elective courses involved, so as to truly provide high-quality curriculum for students. We should make full use of big data thinking in course selection to meet the needs of different students for different courses, so as to promote the all-round development of students. In order to make the effective connection between colleges and students, students can understand the majors, curriculum and advantageous courses of colleges and universities. At the same time, colleges and universities can understand the specific needs of students and really provide targeted courses for students, which requires the use of big data for analysis, induction and summary, so as to provide students with courses with professional characteristics and meet their own development needs in combination with teaching tasks. Therefore, in order to ensure that students can really choose appropriate and professional courses in the course selection process, a big data analysis system should be established to collect, summarize and summarize the data of each course, so as to understand the development of the course and students' satisfaction with the course. By using big data thinking to make the school curriculum meet the development needs of students and adapt to the all-round development of students. For example, students choosing courses is like watching movies on video websites. Once the video websites provide several movies for each person and others are watching them, this behavior belongs to the cooperative screening of big data. In the curriculum and course selection settings of colleges and universities, we can make full use of this big data thinking.

Second, use big data thinking to carry out academic early warning for students. In order to truly cultivate excellent talents, colleges and universities must accurately grasp the learning situation of students. Only when students complete the curriculum and pass the examination of relevant courses can they obtain the graduation certificate and degree certificate [11]. At the same time, students should also know whether they have met the graduation standards in the process of learning, which requires the use of big data thinking for academic early warning. Academic early warning system can not only let students grasp their academic completion, but also let students understand the gap between themselves and graduation requirements. At present, many colleges and universities have adopted the management mode of academic early warning. Therefore, in order to improve the actual effect of academic early warning management, we must use the thinking of big data to innovate and optimize the efficiency of the current early warning management. By using the relevant data of big data, students can grasp their academic status in time. And it is conducive to help students find their own shortcomings and constantly improve, so as to improve their comprehensive quality.

Today, with the continuous development of science and technology, the management of college students has developed in the direction of informatization, but affected by many factors, many work systems have low timeliness and have not been integrated, which is bound to lead to the lag of information exchange, that is, they can only rely on a third party for information exchange; Data transmission is no exception. It needs the support of e-mail and other attachments. Therefore, it can be seen that the transmission efficiency is low. At this stage, the role of student management is becoming more and more obvious, involving many departments. These system barriers greatly increase the difficulty of data audit. During data transmission, counselors need to play the role of system relay identity. It should be noted that the effectiveness of data transmission is closely related to the amount of data requests and information types. Generally speaking, The information category is cumbersome or has large-scale data requests, the timeliness of data transmission is low, and the workload of counselors is increased.

College student management is a relatively complex work, involving many aspects, and different contents correspond to different departments. In other words, each department needs to undertake different work, and student information also needs to be classified in combination with reality, that is, each department stores and manages student information based on classification. For example, educational administration information should be stored in the school education department Logistics information should be stored in the logistics department, etc. Due to the lack of good communication and interaction between departments, the mobility of student information is not strong. If you want to use the student information in charge of other departments, you need to submit an application. You can use the required information only after obtaining the review and permission of the head of the Department [12]. The coordination efficiency is generally low. With the continuous progress and development of science and technology, some colleges and universities have gradually realized the importance of student management informatization and began to try student information management. However, affected by the language factors developed by developers, various subsystems are difficult to integrate and operate, reflecting a certain degree of independence. Whether it is data export, data format processing or data import, they are relatively closed, It is prone to errors, which affects the collaborative efficiency.

There is a certain neglect of data integration. The information involved in student management is mostly dynamic information, such as students' classroom performance, students' original information, etc. whether from the perspective of type or structure, these data have a certain complexity and are related to many departments. Compared with other data, grass-roots data has certain particularity, which is inconsistent with the required form. Errors are easy to occur in the process of data collection, which increases the difficulty of data collection; Affected by useless data or data format characteristics, the data summary results will also change to some extent, and the accuracy needs to be improved.

4 Simulation Analysis

Decision tree is a tree structure that can automatically classify data. It is the knowledge representation of tree structure and can be directly transformed into decision rules. It can be regarded as the prediction model of a tree. The root node of the tree is the whole data set space. Each sub node is a splitting problem. It is a test of a single variable, The test divides the data set space into two or more blocks, and each leaf node is a data partition with classification. Decision tree can also be interpreted as a special form of rule set, which is characterized by the hierarchical organization of rules. Decision tree algorithm is mainly used to learn the learning method with discrete variables as attribute types. Continuous variables must be discretized to be learned.

Student management system, as its name implies, is an information system that manages and analyzes all kinds of information of students [13]. It can effectively support the decision-making and control of educational administrators, as well as the use and control of information by students themselves. Student management is different from the general management information system. It is the specific application of management information system in the field of student management. However, student management can learn from the theory, method and development process of management information system and apply it to specific practice. Different from the management information system, student management has its own important characteristics:

College student management is a very specific and detailed work. All links are interrelated. It not only has a large workload, but also is very complex and involves a wide range. For example, in the process of student status management, performance management and curriculum setting, it is necessary to repeat the operation of a large number of forms and make accurate statistics on these forms. Moreover, with the expansion of the scale of colleges and universities, there are more and more students in Colleges and universities, followed by a large amount of student information and a large number of data statistics. The cumulative information of each student is also amazing, which brings great pressure to the student information management of colleges and universities.

Load the CPU data set in the preprocessing tab, select the linearregression classifier in the classification tab, and select the 10 fold cross test. By default, click start to generate the regression equation and test results, as shown in Fig. 4 below.

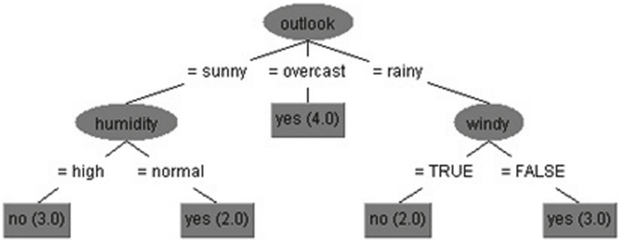

Fig. 4. Data set decision tree

Student management system (hereinafter referred to as the system) is a powerful tool for educational information management. It can comprehensively manage students' academic performance, behavior performance, semester comprehensive evaluation, excellence and awards in the whole process from enrollment to graduation. After data accumulation, a large amount of data is collected [14]. In order to find the valuable information hidden behind such a large amount of data, data mining technology is applied to the system to form a data mining subsystem. It mainly analyzes the student information and draws a scientific and reasonable conclusion. As shown in Fig. 5 below, the operation results of student management using Weka are shown.

```
=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9012
Mean absolute error                     41.0886
Root mean squared error                 69.556
Relative absolute error                 42.6943 %
Root relative squared error             43.2421 %
Total Number of Instances              209
```

Fig. 5. Weka operation results

5 Conclusion

This paper studies the application of data mining technology based on Weka in student management. Using Weka software, this paper applies the decision tree algorithm to the analysis of student information management system, selects representative data indicators, and tests the sample data. University managers can use the test results to analyze the management of colleges and universities and student information, and use data analysis to mine valuable information from behind the data, so as to better manage students. The research significance of this paper is to transform student management information into effective, orderly, regular and decision-making knowledge and promote the construction of campus digitization.

References

1. Abdullahi, K., Kadir, A.: Social justice issues in education and management of student development in Nigeria. Contemp. Educ. Res. J. **11**(3), 78–91 (2021)
2. Han, X: Design of counselor student management system based on competency model. In: 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (2021)
3. Razak, F., et al.: Investigating the linkage between service quality and satisfaction in context of student management system: an evidence from Malaysia. J. Phys: Conf. Ser. **1793**(1), 012033 (2021)
4. Fata, M., et al.: Curriculum development management in increasing student excellence research context (Multisite Study at MTsN 1 Trenggalek and MTsN 2 Trenggalek) (2021)
5. Damayanti, D.: Perbandingan akurasi software rapidminer dan weka menggunakan algoritma K-Nearest Neighbor (K-NN) (2021)
6. Geyik, B., Erensoy, K., Kocyigit, E.: Detection of phishing websites from URLs by using classification techniques on WEKA. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT) (2021)
7. Ahmed, A., et al.: An approach to detect cyber attack on server-side application by using data mining techniques and evolutionary algorithms (2021)
8. Wiweka, N.A., Mahadewi, L., Suwatra, I.: Pengembangan multimedia flashcard bilingual terhadap mata pelajaran bahasa bali kelas I SD (2021)
9. Suman, Mittal, M.P.: A comparative performance analysis of classification algorithms using weka tool of data mining techniques (2021)
10. Dnler, Z.B., Ahn, C.B.: WEKA ortamn kullanarak derin renme ile kimlik hrsz web sitelerinin tahmini. Eur. J. Sci. Technol. (2021)
11. Jalal, A.A., Ali, B.H.: Text documents clustering using data mining techniques. Int. J. Electr. Comput. Eng. **11**(1), 664–670 (2021)
12. Akyürek, M.B.: Time management skills of university students. Yuksekogretim Dergisi **11**(1), 139–147 (2021)
13. Moon, M., Oh, E.G., Baek, W., Kim, Y.M.: Effects of nurse-led pain management interventions for patients with total knee/hip replacement. Pain Manag. Nurs. **22**(2), 111–120 (2021). https://doi.org/10.1016/j.pmn.2020.11.005
14. Efthymiou, L., Zarifis, A.: Modeling students' voice for enhanced quality in online management education. Int. J. Manag. Educ. **19**(2), 100464 (2021)