Application of Classification Mining Technology Based on Decision Tree in Student Resource Management

Hongting Li(✉)

Wenhua College, Wuhan 430074, China
lhh1028@163.com

Abstract. This paper discusses mining method based on decision tree in student resource management. With the increase of college enrollment, the number of students is also increasing. Oral teaching management produces a lot of data every day, and the existing teaching management system is becoming more and more problematic. With the wide application of database management system, the ability of data collection has been greatly improved and a large amount of data has been accumulated. Behind these data are very important and valuable information. The establishment of data mining technology is to analyze these data at a higher level and obtain the potential information of future operation and life. Data mining is a process of extracting hidden, unknown but potentially useful information and knowledge from a large number of incomplete, noisy, fuzzy and random fact data. How to transform these data information into knowledge representation, reasonably use these information to serve teaching management, scientifically guide teaching and improve teaching management level is an urgent topic for us to study. Data mining technology is a feasible and effective method to solve this problem.

Keywords: Decision tree · Excavation technology · Student resource management

1 Introduction

At present, teaching resources show a large number, many types and miscellaneous forms. It is a problem to be solved to realize the effective organization and management of these resources, enable teachers to quickly search and browse the required information in enable teachers to publish courseware conveniently and quickly.

In addition, the school-based resource management system mainly serves for school education and teaching, so as to improve teaching efficiency and promote the all-round development of students [1]. Therefore, the system is also required to realize certain user interaction and become teachers to communicate.

The school-based resource management system is to facilitate the administrator to manage the teaching resources. Running the application software can be used with

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 149–160, 2023.
https://doi.org/10.1007/978-3-031-23950-2_17

less system cost. When it is put into operation, it can save a lot of human and material resources for the school's teaching resource management. The benefit brought by the system is far greater than the development cost of the system software, which is convenient, effective and simple.

After careful analysis, the designed interface of the school-based resource management system is concise and clear, taking into account some habits of people when using as much as possible, making the operation simple. After each operation, it will respond to the user, with flexible and reliable data reading, complete functions and easy use.

Data extraction methods have obvious advantages in analyzing large amounts of data. Analysis methods based on data extraction have been widely used in finance, insurance, telecommunications and other major industries. With the development of data mining technology and the expansion of its application scope, some university scholars began to study university data mining technology in the field of teaching management. For example, using relevant rules to study the hierarchical structure of relevant professional courses and the relationship between CET-6 management and CET-4 management, the research results have made an important contribution to improving the level of school education management.

The application of data collection technology in family education started late. In particular, the data mining technology used by students for data analysis and processing is rarely reported in the literature. The traditional student management and analysis methods generally only include data backup, retrieval, statistics and sorting [2]. Examination management data contains a lot of useful information. If it is not used, it is easy to waste data. At present, the number of students in a university is as few as thousands or as many as tens of thousands, and the examination management data can easily reach millions or even tens of millions, which makes the traditional management analysis methods more difficult to meet the needs of evaluation.

As the support of decision-making process, data extraction is a detailed data analysis method. It will be very useful to apply data extraction technology to management evaluation. We can comprehensively analyze the internal relationship between audit management. For example, data collection tools can answer similar questions, such as "factors affecting student management", which cannot be answered by traditional evaluation methods.

At present, digital campus has become an important goal of school information development in China. In the process of digital campus construction, a lot of information has been accumulated, and the amount of data in the database is increasing. If we still use simple statistical methods, it is obvious that we cannot find the connection between data and rules. Data extraction methods can obtain useful information from these massive data. In order to solve this problem, data mining technology came into being [3].

Data mining method is used to analyze and deal with students' learning, get students' evaluation results, and correct students' bad behavior in time. Give full play to the examination effect, comprehensively evaluate the teaching quality, timely feed back the teaching results, and communicate the teaching information.

2 Related Work

2.1 Decision Tree Algorithm

Decision tree is a data classification algorithm based on predictive variables As a prediction model, decision tree draws the conclusion of target variables through the analysis of prediction variables. The decision tree is constructed in the same way as the flow chart. All internal nodes are test attributes, and all nodes are test results. One advantage of decision trees is that they are relatively simple and easy to describe. To create a decision tree, select an input variable (target and multiple presets), and then create a backlight in the detector until a valid structure is created. The decision tree consists of nodes, roots, leaves (sometimes called terminal nodes) and branches. Figure 1 below shows the decision tree flowchart structure.

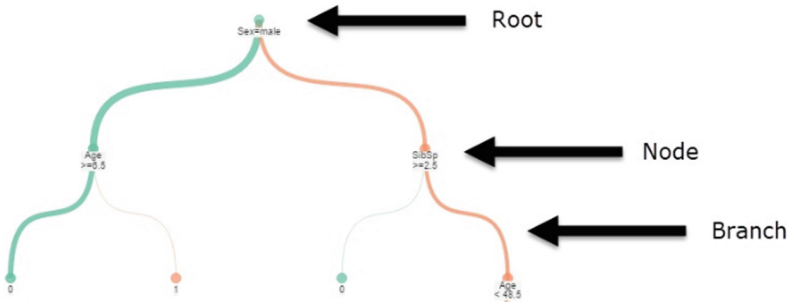


Fig. 1. Decision tree flowchart structure

Classification is a very important data extraction method According to the adopted classification model, data classification methods mainly include decision tree generalization, Bayesian classification and Bayesian network, Decision tree classification algorithm is one of the most widely used algorithms in generalized reasoning [4].

Each node defines the attributes it represents. On the basis of the decision results, different branches of input nodes and leaf nodes represent the classification results.

The top of the node is the root and the end is the leaf. The whole paper is used to predict the output of Y variables in the classification section The internal (non terminal) node specifies the decision point of the data partition. Branches are the results of each test (node).

Decision trees can be used for classification or digital data. If you want to predict which variables will be classified, create a classification tree. If the target variable is continuous, a regression tree is created, but it is different in some aspects. The most obvious is the partition definition (the variable threshold of the segmented data).

There are many algorithms to build decision trees. Here we only discuss two kinds:

1. Classification and regression tree - use Gini index (classification) as the measure.
2. ID3 (iterative dichotomizer 3) - use the direct function and information gain as the measurement, as shown in Fig. 2 below.

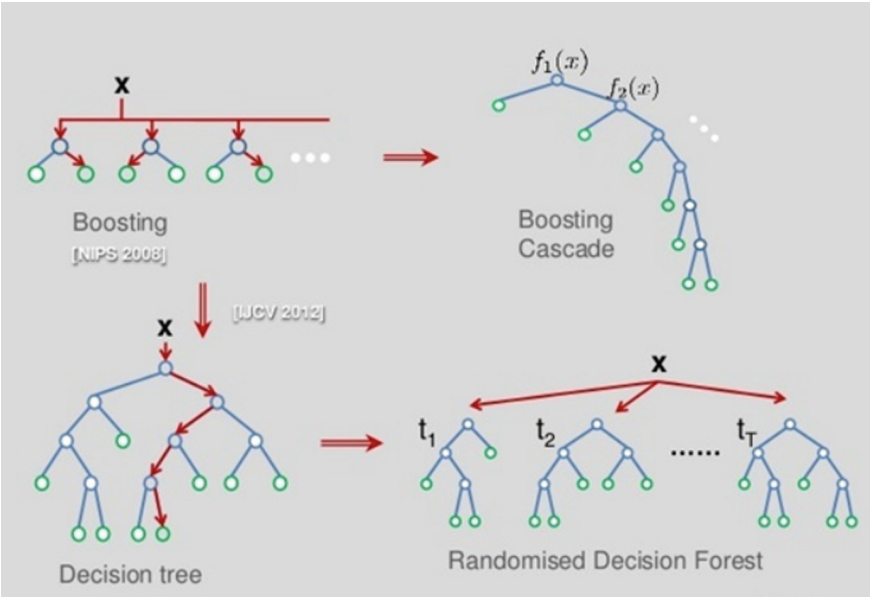


Fig. 2. Direct function and information gain

Random forest is a model that can achieve very, very good classification and regression results in machine learning. We often see data competitions like datacast, which also use random forest very much. We can also get better results by directly classifying or regressing random forests.

2.2 Data Mining Processing

The data source used for data mining must be real and large, and may be incomplete and include some interfering data items. The discovered information and knowledge must be interesting and useful to users. Generally speaking, the result of data mining is not required to be completely accurate knowledge, but to find a general trend.

Data mining can be simply understood as the process of discovering useful knowledge through the operation of a large amount of data.

The knowledge discovery of data mining is not to discover universal truth, to discover theorems and pure mathematical formulas, nor to prove machine theorems. In fact, has specific preconditions and constraints, faces specific fields, and can be easily understood by users [5]. It is best to express the discovered results in natural language. Figure 3 is an example of several common correlations.

Data mining can not only establish descriptive (retrospective) models from historical data, but also establish predictive models, which provides a powerful solution tool for us to extract useful information from large-scale databases. Data mining knowledge. The knowledge obtained through data mining is "explicit", which can not only be understood, but also easy to store and apply. Therefore, it has been widely valued as soon as it appears [6]. A large number of different situations of known results can be stored in the computer,

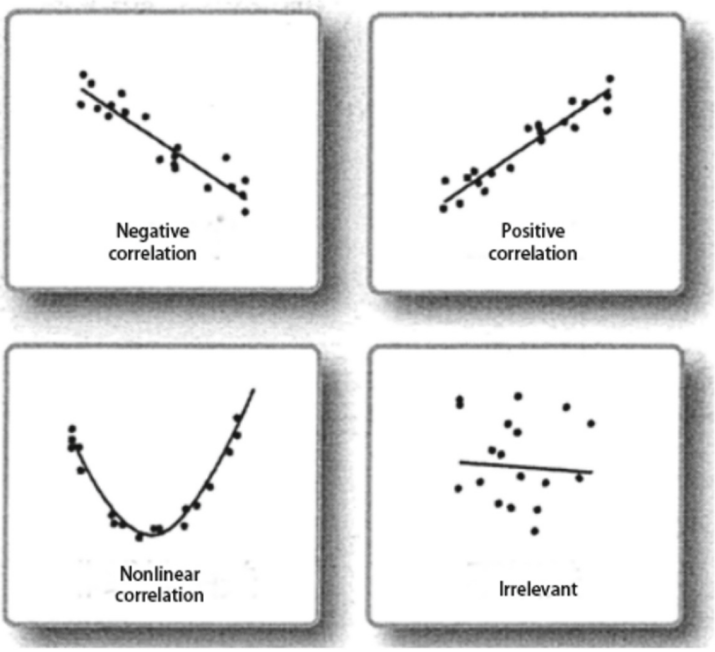


Fig. 3. Examples of dependencies

and then the data mining tools can search for gold from these information, extract the information that can produce the model, and express the model in a way that is easy to understand, such as graph, table, formula and so on.

Data mining is a young interdisciplinary field. It integrates the theories and technologies. It is the result of the evolution of information technology.

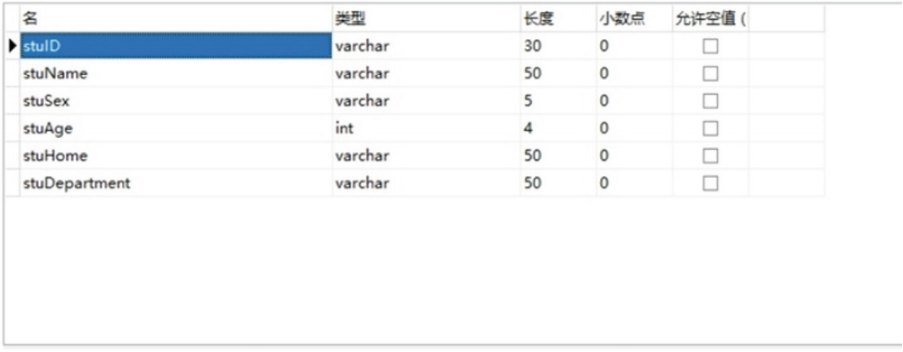
A typical data collection system consists of the following parts:

1. Database and data warehouse. A database, data warehouse, or other repository is a database or database, data warehouse, spreadsheet, or other type of repository. Data processing and synthesis.
2. Data retrieval machine is the most important part of data extraction system. It includes functional modules such as symptom, correlation, classification, cluster analysis, evolution and variance analysis.
3. Using models to estimate and measure interest rates is helpful to design data acquisition modules, with the focus on extracting more meaningful models.
4. Graphical interface is the bridge between users and data retrieval system It allows users to interact with the system, define queries or tasks related to data extraction, and provide information that helps to find the focus, which depends on the intermediate data retrieval results of data retrieval.

3 Application of Classification Mining Technology Based on Decision Tree in Student Resource Management

The especially the education student management system, helps the training institutions to realize information management and improve management efficiency from the aspects of student file management, registration management, class scheduling management, class hour and class consumption management, growth files, classroom comments, home school interaction and so on. Real database systems often store a large number of diverse data, which expand rapidly. Therefore, usually, the data we collect often contain a lot of noise data, and even some incomplete and inconsistent data. Therefore, the data objects involved in data mining must be preprocessed to ensure the integrity and consistency of data, which is an important step in the process of data mining. It helps to improve the quality of data mining objects.

Data integration is to combine. This study uses database technology. The collected database table files "student basic information table", "teacher basic information table" and "student management information table" are integrated and merged to generate "initial table of student management data analysis" by using the connection operation of tables in the database. Since our school stipulates that the management entry person of the course is the teacher itself, the "teacher basic information table" and the "student management information table" are connected through the "teacher job number" field and the "management entry person" field, and the generated data table is connected with the "student basic information table" through the "student number" field. The "initial table of student management data analysis" generated after this connection includes the following fields: student number, student name, major, province of origin, enrollment management, job number, teacher name, teacher gender, teacher title, educational level, course code, daily management, examination management, comprehensive management, management input person, remarks and other attributes, The number of records is 1157 lines. As shown in Fig. 4 below.



名	类型	长度	小数点	允许空值 (
stuID	varchar	30	0	☐
stuName	varchar	50	0	☐
stuSex	varchar	5	0	☐
stuAge	int	4	0	☐
stuHome	varchar	50	0	☐
stuDepartment	varchar	50	0	☐

Fig. 4. Design data sheet

The "initial table of student management data analysis" includes many attributes, such as student number, student name, student gender, major, student source Province,

enrollment management, job number, teacher name, teacher gender, teacher title, educational level, course code, daily management, examination management, comprehensive management, management entry person, remarks and so on. In the actual data classification process, we can only select some attributes as classification objects [7]. Therefore, we need to analyze and summarize the attributes in the table, remove the attributes that have nothing to do with the analysis of management data or have poor correlation, and select the attribute values related to management data as the research object. In order to facilitate the establishment of the decision tree model, this paper selects the three attributes of "student gender, peacetime management and teacher title" which are closely related to the management attributes as the basis for establishing classification decision tree model, and generates the "student management data mining table".

1. Determine the object and goal of data classification. The objects of data classification often need to be clearly defined according to the actual situation of data mining database. Due to the unpredictability of data classification results. In addition, is very important for the establishment of data classification objects and even the generation of the whole classification results.
2. Data collection. Data acquisition covers a wide range and has a variety of ways. For example, we can collect student achievement data and teachers' basic information data; We can obtain data from some existing databases, or collect and summarize data through questionnaires. In short, data acquisition is a time-consuming and laborious work.
3. Data preprocessing. This stage usually includes four links: data cleaning, data integration, data transformation and data summarization. The purpose of data processing is to ensure the integrity and consistency of data in the database.
4. The final result of data mining is to establish a classification model. First, according to the test objectives, formulate appropriate data extraction methods, such as the classification method used in this work. Secondly, like C45 decision tree algorithm, we should design a classification mining algorithm according to the actual location of the test data set, classify the student performance data, and finally establish a "student decision tree model".
5. Formulate classification rules. This will make the classification rules of mining results consistent with the previous data classification. This means rules for extracting symbols from visual classification models. It is worth noting that the decision tree generated in advance is often pruned before generating the classification rules, otherwise the classification rules may be meaningless or inconsistent with the objective law due to the over training of the decision tree by the classification algorithm [8].
6. Application of knowledge. The so-called knowledge here is the classification rules generated in the previous link. For the application of classification rules, we can not only directly apply them to daily teaching activities, but also predict new data. For example, this paper makes full use of the knowledge generated by the experimental results, obtains the conclusions and methods to guide the teaching, and systematically realizes the data prediction function.

4 System Design

Because the construction of educational resources has the characteristics of regional universality, technological complexity and cultural diversity, it is difficult to share a large number of educational resources and communicate with different educational systems. The key to solve this problem lies in the standardization of educational information. At present, many countries and regions in the world have established organizations specializing in the standardization of educational information, and formulated a series of educational information standards, involving learner related, content related, data and metadata, management systems and application software. Among them, the metadata standard of educational resources is the basis for effectively describing educational information resources and realizing resource discovery and exchange.

To realize the co construction and sharing of educational resources and build a distributed and shared educational resource management system is an active project all over the world. Metadata with good interoperability and based on resource discovery and retrieval is the basis and key to effectively realize the description, retrieval and utilization of these resources. The co construction and sharing of educational resources have been carried out in countries all over the world. The United States is one of the countries with the most abundant educational resources in the world. The U.S. government, universities and non-governmental organizations have invested a lot of funds and research forces in the construction and use of network resources. Among them, representative ones are Eric (American Educational Resource Information Center), free (Federal quality educational resources), Ohio link (Ohio Library and information network) and Ceres (Educational Resource Center project), etc. In the United States, it is very common for schools to jointly build and share educational resources, such as the MIT open courseware project, which has attracted the attention of many countries and universities around the world.

The overall design includes the following key designs: the design of computer configuration, the design of system module structure, database and various files. When we carry out this task, we need to separate all parts of the software, and design different modules carefully to make it have its own functions. Software decomposition is also called divide and conquer [9]. First separate the functions of the software, and then integrate each module to make them a whole. This design should be based on the following principles as far as possible;

1. Practical: designed according to the actual needs;
2. Easy maintenance: adopt a good model for design, so that the system can be easily maintained in the later stage, so that the school does not need to invest too much money;
3. Convenient operation: the designed system needs to be convenient for users according to the situation of the school, and can not make the operation of the designed products troublesome and wasteful;
4. Safety: the designed products are closely related to the privacy of students, so the safety factor must be particularly high;
5. Reasonable function: the system we designed must ensure that each function can be used by users, so we need to make the system both beautiful and practical when designing;

6. Design time: we should not only ensure the quality of the system, but also ensure that it can be developed within a reasonable time.

SSM is a combination of three frameworks (Spring + spring MVC + mybatis). The framework is a relationship mapping and a solution of reusable software architecture. SSM is an efficient development framework commonly used at present. Because I have done practical training before, the teacher can directly design a similar system with only Maven. Relatively speaking, SSM seems to use more frameworks, but in the design process, the development efficiency of SSM is really very efficient. The logic of the background is not as complex as maven, so I learned the SSM framework to design the system, as shown in Fig. 5. The design works are relatively simple and can be used as a basic reference.

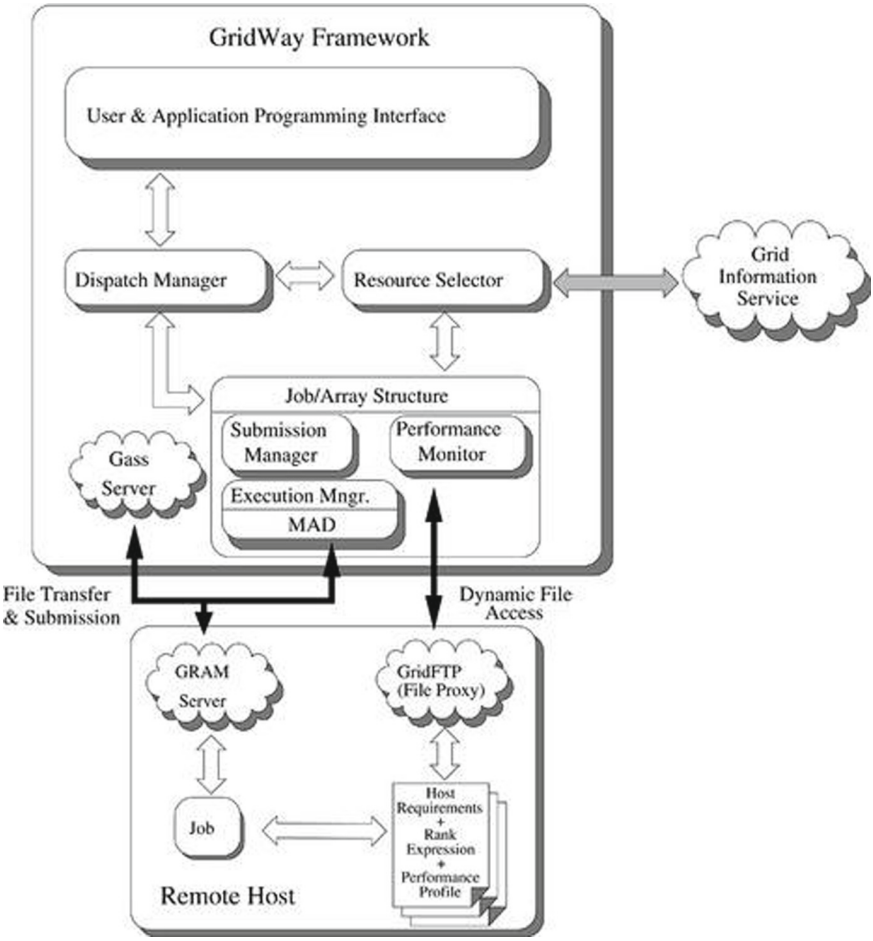


Fig. 5. Student resource management system

Integrating the three, SSM uses spring MVC as the controller of hierarchical management. Springloc is responsible for managing component resources, and spring AOP is responsible for aspect programming management. Mybatis acts as the database access persistence layer, realizes the data interaction between Dao in spring and data control system, and is responsible for database transaction management. SSM framework is suitable for building various large-scale application systems, which can be applied to school educational administration system management, student dormitory management, enterprise management, government department management and so on [10].

Analyze the system function and performance requirements to realize that different user systems have different operation permissions. After the user types of login system are generally divided into student users and administrator users, enrich the management content of student information, and set modules on the system page to display the diversified information of students. After distinguishing user types in the login interface, different modules are displayed in the system menu bar, such as grade, class, students' basic personal information, management, course selection and other modules.

After implementing the three-tier structure, the business layer through a unified interface. The company processes queries according to its logical database rules and returns data from the database to user classes. Therefore, the user interface cannot understand the database structure. Maintain the interface with the business phase. This method can improve the security of the database to a certain extent and reduce the requirements of user interface developers. Because there is no need to run the database. The JavaBean above returns data in object format. This allows you to specify a class in which you can access read-only data and therefore encapsulate the data to further improve data security. As shown in Fig. 6 below, the data development code.

```
Class.forName("com.mysql.jdbc.Driver").newInstance();

String url="jdbc:mysql://localhost:3306/xscj_database";

conn=DriverManager.getConnection(url,"root","111111");

Statement stmt = conn.createStatement();

}catch(Exception ee){

System.out.println("connect db error:"+ee.getMessage());
```

Fig. 6. Data development code

Since learning management is a kind of information, the management process is the process of obtaining, storing, transmitting, processing and outputting this kind of information. However, many colleges and universities ended their work after marking, management publishing, reporting, classified registration and storage, without in-depth analysis from a large number of management and mining information conducive to

teaching work, Taking these as the end of the work is a waste of teaching information resources [11]. The reason can be attributed to the tedious work, scattered data, traditional manual operation, high labor intensity and low efficiency, which is not convenient for query, classification and summary, and can not provide all kinds of information in time and accurately. The traditional student learning management analysis is nothing more than to get the mean, variance, difference significance test, reliability and so on. It is often considered based on the teaching itself. These need further analysis to draw a conclusion for education managers to make corresponding decisions, but these information can not be obtained by using the traditional student management analysis methods, real reasons affecting students' learning management, we can formulate corresponding measures to improve the quality. The studying management is to find out the factors affecting the regularity of management change through the disclosure of the law, and on this basis, "suit the remedy to the case" [12]. At the same time, restrain negative factors on students and guide them to study actively and develop healthily. Processing of large-scale data. It can extract the useful information hidden behind the data. It is being used in more and more fields and has achieved good results. It is very beneficial. It can find examination management and various factors through the comprehensive analysis of relevant data. For example, through the analysis of student related data, data mining tools can answer similar questions such as "which factors may have an impact on student management", This is not available in traditional evaluation methods.

5 Conclusion

With the valuable information emerges from a large number of data collected by data mining technology. At present, data extraction technology has been widely used in biomedicine, financial data analysis, e-commerce, telecommunications and other fields. At the same time, there are few examples of successful application of data extraction technology in university education system. In view of the general trend of educational informatization, data mining technology is used to obtain the information of educational activities, so as to lay a foundation for teachers' educational activities in the future, improve teaching methods and improve the quality of education. This paper analyzes the design of student information management system. The design of the system includes the overall design and creation of the database and the data table structure. It describes the company.

References

1. Letina, A., Dikovic, M.: Student teachers' classroom management orientations and their beliefs on effective teaching behaviour, sodobna pedagogika. J. Contemp. Educ. Res. **72**(138), 128–143 (2021)
2. Thamrin, R.M., Andriani, R.: Design web-based registration and data management of student thesis information system. SISFOTENIKA **11**(1), 101 (2021)
3. Liu, Z., Cai, Y.: Construction management of student affairs center guided by the student-based idea: taking the business school of beijing institute of technology zhuhai institute as an example. In: 1st International Conference on Education: Current Issues and Digital Technologies (ICECIDT 2021) (2021)

4. Mao, L., Zhang, W.: Analysis of entrepreneurship education in colleges and based on improved decision tree algorithm and fuzzy mathematics. J. Intell. Fuzzy Syst. **40**(2), 2095–2107 (2021)
5. Dinesh, T.: Higher classification of fake political news using decision tree algorithm over naive bayes algorithm. Revista Gestão Inovação e Tecnologias **11**(2), 1084–1096 (2021)
6. Hardiani, T.: Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta. Int. J. Health Sci. Technol. **2**(1) (2021)
7. Arigi, A.M., Park, G., Kim, J.: An approach to analyze diagnosis errors in advanced main control room operations using the cause-based decision tree method. Energies 14 (2021)
8. Rahmawati, A., Yan, R., Riana, D.: Deteksi defect coffee pada citra tunggal green beans menggunakan metode ensamble decision tree. Technol. Commun. **20**(2), 198–209 (2021)
9. Djordjevic, D., Cockalo, D., Bogetic, S., et al.: Predicting entrepreneurial intentions among the youth in serbia with a classification decision tree model with the QUEST algorithm. Mathematics 9 (2021)
10. Gatwood, J., Hohmeier, K., Kocak, M., et al.: Acceptance of productivity software as a course management and collaboration tool among student pharmacists. Currents in Pharmacy Teach. Learn. **13**(4) (2021)
11. Boomars, R., et al.: Optimizing airway management and ventilation during prehospital advanced life support in out-of-hospital cardiac arrest: a narrative review. Best Practice Res. Clin. Anaesthesiol. **35**(1), 67–82 (2021)
12. Scott, A.T., Dpm, C., Dpm, T., et al.: Review of achilles tendon reattachment using double-row knotted and knotless techniques in the management of insertional achilles tendinopathy (2021)