

PE Teaching Activities in Colleges and Universities Based on Decision Tree

<https://doi.org/10.3991/ijet.v13i08.8693>

Jian Chen

Hunan Vocational College of Science and Technology, Hunan, China
jchen096581@21cn.com

Abstract—To improve the efficiency of student's sports performance management and the quality of physical education teaching, the overall structure of the sports performance management system based on decision tree was designed, and a systematic database design plan was proposed. Then, ASP.NET technology was used. Combined with SQL Server database, on the Visual Studio platform, the implementation process of sports test type, test items and score management was given through the flow chart of the program. Moreover, the decision tree of the relationship between the sports performance and the students' employment was generated. It was proved that the efficiency of the sports teachers' entry and statistical results were improved through the construction of this system. It is concluded that the sports performance and the sports teaching based on rule are better developed through the system of data mining.

Keywords—Decision tree, data mining, colleges and universities, sports achievement

1 Introduction

Today, information technology has become an indispensable part of life in the information age. With the deepening and widespread application of information technology, a lot of data that need to be stored and reused have been generated, which leads to a new era of big data. In the face of such huge data, data mining technology has emerged. The technology is to find the link between data and data and the hidden secrets of the data. Data mining in China started late but progressed rapidly. Some big enterprises and high-tech enterprises have done a lot of research and preliminary commercial use, and have made many achievements.

As an important part of national education, the construction of information technology in Colleges and universities cannot be ignored. However, the application of information technology is not widely used in school physical education. The large number of physical education teachers are faced with the actual problem of the large workload in the annual sports examination stage. The application of decision tree and data mining technology in physical education is discussed in combination with the practice of physical education in school. The purpose of this study is to realize a

sports performance management system based on data mining, to serve the physical education in schools and to improve the quality of physical education.

2 Literature review

Since Von Neumann defined the internal data of computers as binary representations, computer scientists have been searching for the best way to organize data in computer software. The Hasegawa and others as stated in [1] further studied the atom/ion movement controlled devices for beyond von-neumann computers. In the 70s of last century, Edgar Codd as stated in [2], a scientist at IBM, invented the relational database. He completely separated the program from the data, and solved the problems of complex organization structure, hierarchical system in the past database, the adhesion between programs and data, and the lack of independence. It has become a significant milestone in the history of computer software development. Big data brings a lot of data. Stuart as stated in [3] proposed the data revolution. The main content mainly concludes big data, open data, data infrastructures and their consequences. In these data, we find mutual connections to help users find the key points of development and predict future trends. Thus, data mining technology has emerged. Wang and others as stated in [4] discussed the application of data mining technology and made great improvement in some fields.

The first application of decision tree technology is in the concept learning system, which lays the foundation for many decision tree learning algorithms that appear later. Accordingly, Schmitt as stated in [5] proposed the decision tree algorithm. Ramathan and others as stated in [6] predicted students' performance using modified id3 algorithm. In addition, to better design the PE system, the famous ID3 algorithm at the end of the 70s of last century was used. Niu and others as stated in [7] predicted the protein solubility by integrating chaos games representation and entropy in information theory. The concept of entropy in information theory is applied in this algorithm for the first time. Furthermore, the division attributes are determined by using the information gain to measure the discriminant ability of the decision attribute classification. Prasad and others as stated in [8] pointed out the SLIQ: performance enhancement in sliq decision tree algorithm. In addition, a fast and extensible classification algorithm SLIQ at the late 90s of the last century is applied. In addition, Lu and Yan as stated in [9] applied the medical data mining of unmixed interval inductive based SPRINT algorithm. Therefore, an extensible and parallel inductive decision tree algorithm - SPRINT algorithm is proposed, which is comparable to SLIQ. Moreover, the PUBLIC algorithm in the view of "building trees and pruning together" is put forward. Later, Gao and others as stated in [10] proposed the rain forest algorithm. Therefore, the rain forest algorithm that can maximize the use of memory resources is applied.

In summary, there are still some loopholes in the classroom design of sports teaching activities, such as the daily teaching content and the group cooperation project. In view of the shortage of classroom design, a new design concept based on decision tree

is proposed. This method makes up the shortage of the current sports teaching management system and can improve the quality of physical education better.

3 System design

3.1 The overall architecture of the system

The overall technical architecture of the system is shown in figure 1. As shown in the figure, the system is divided into three layers from the top to the bottom. The system is divided into three layers from the top to the bottom. The top layer includes a control page for the users of sports teachers, students, and system administrators. The middle layer is the business logic layer, which is responsible for obtaining data from the page, making logical judgment and interacting with the data at the bottom. The bottom layer is the data access layer, which is responsible for the interaction with the database.

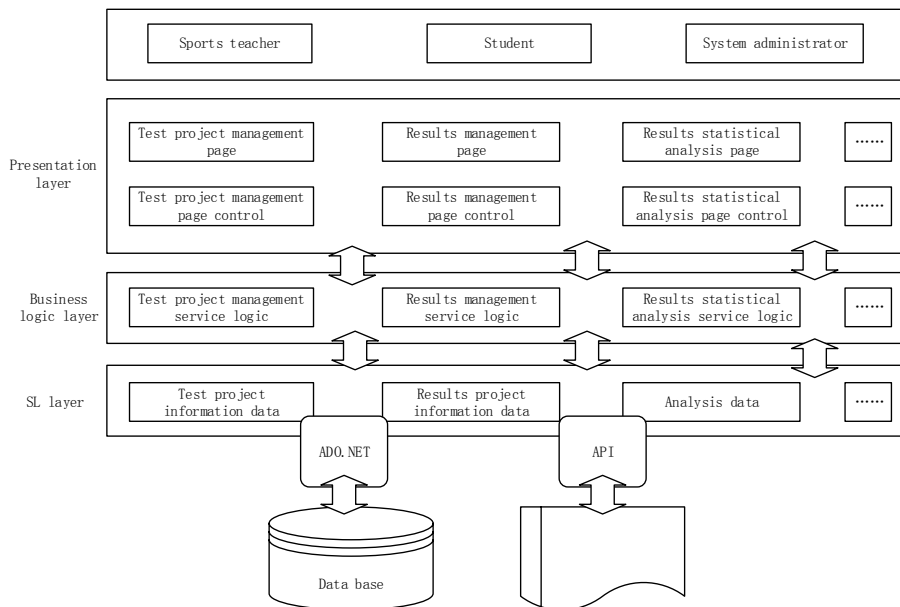


Fig. 1. System architecture

The development of the system uses the Net three-layer architecture. Each business function is divided into three layers. The top layer is the user layer, which provides a visible, operable, and friendly functional interface for the system users. Users do not need to care about how these features are implemented. The middle layer I is the business logic layer. This layer mainly obtains user's input data at the interface, makes logical judgement on business. Through calculation, the data in the database are feed-back to the user interface for users to browse to a certain form. It can be said that the

business logic layer mainly explains how the function is implemented. The bottom layer is the data access layer. The layer first obtains the data processing results of the middle layer, and then performs the operation of adding, deleting, modifying and checking for database. Finally, the database operation results are fed back to the business logic layer for processing in the middle layer.

The greatest advantage of the Net three-layer architecture is: Each layer is separated from each other and minimizes the coupling between functional modules. This is in line with the idea of "low coupling and high cohesion" in software engineering. From the perspective of developers, this method does not need to modify all of the code when the business needs change, only the business logic needs to be adjusted. Therefore, it improves the efficiency of the development of the program and reduces the cost of development.

3.2 Database design

Database design refers to the process of creating database and designing detailed data structure in a specific database management system after system analysis. In the process of software development, database design is the core of the system design, and it is the soul of any information system. To run the program running of a complex business system, it is more complex when the database is designed. Therefore, when designing database, we must follow the principle of "gradual progress". The specific process is to plan the data entity and analyze the entity relationship, and then to design database table structure. In software development, a standardized database design is usually implemented in three steps: database concept design, database logic design and database physical design.

The main work of developing software is to design a database and develop an application built on the database. In the use of software, users are concerned only with the implementation of the software function without understanding the implementation of the data level. Therefore, database design is not important for a lot of people. However, on the contrary, the database is the soul of the whole software system. Database design is to develop the related data table structure to meet the needs of data storage and processing. In the design of the database, we should follow the principles of safety, specification, accuracy, consistency and efficiency, and ensure that the database can better serve the software system.

The target system includes the user information entity, the student information entity, the test type entity, the test item entity, the test score entity, and the achievement information entity. Figure 2 shows the user information entity. The attributes of the user information entity include the user name, the login password and the user type.

Figure 3 shows the student information entity. The attributes of the student information entity include student name, school number and login password.

When designing standardized database logic, we should also consider the proper destruction of standard rules, that is, anti-normalization design. Query speed is accelerated by reducing the number of indexes, tables, and the number of connection operations. The commonly used anti - Specification technologies include adding redundant columns, increasing derived columns and reorganizing tables.

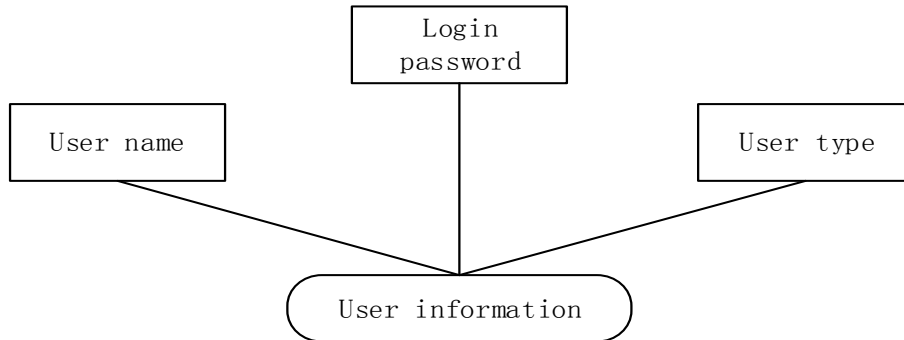


Fig. 2. E-R diagram of user information entity

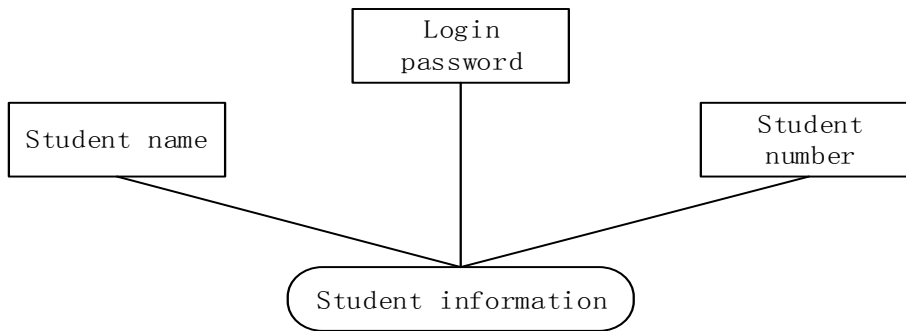


Fig. 3. E-R diagram of student information entity

Then, the logical design of the database is carried out. The database logic design is the first half of the whole database design, including the required entities and relationships, and the entity normalization and so on. The logical design of the database determines the overall performance of the database and its applications and the location of the database. If the database logic is not well designed, all the tuning methods are limited to improve the performance of the database. To further improve the method of database design, the normalization theory of database must be observed. The theory of standardization provides theoretical guidance and tools for the logical design of database. It reduces the data redundancy and saves storage space. At the same time, the speed of increasing, deleting and changing is speeded up.

Database logic design is the second step of database design. The premise of logic design is that we have completed the conceptual design. The design content is to transform the conceptual design into a logical structure consistent with the database management system. Logical database design can be divided into three steps: First, the concept structure is transformed into the corresponding data model, such as the relational model, the mesh model and the hierarchical model. Secondly, the data model is transformed into the data model that the corresponding database management system can support. Finally, the transformed model is optimized.

First, we must select the most suitable data model to express and describe the conceptual structure, then choose the appropriate database management system (DBMS). The E-R diagram is transformed into a relational model, that is, the entity in the E-R diagram, the relationship between the entity and the entity attribute are transformed into the corresponding relational schema. In the conversion, we should follow some basic principles, and the basic principle is that an entity can only be transformed into a relational model. After transformation, the property of the entity is the relation attribute of the relational schema, and the entity code is the code of the relationship. After the transformation of the relationship model, it is necessary to optimize the data model. The optimization of data model is to determine the dependence between data and data and eliminate redundant connections in relational models. The specific way is to merge data items of the same attribute. When optimizing, the standard of 3NF is usually followed. That is: Every row in the table should not have repeated data. Each record should have a primary key that can be uniquely distinguished. The table cannot store non-key data that depends on other keys.

According to the above conversion principle, the E-R map of the sports performance management system is converted. The following 6 relational patterns are obtained: user information (user ID, user name, login password, user type), student information (student ID, student name, login password, test type (test type ID, test type name, test type description, test type weight), test project (test project ID, test project name, test type, test project description, test project weight), test score (test score ID, test item, school number, score), performance information (score ID, school number, score).

The result of this step is the "sprint database". Based on the requirement analysis and the concept and logic design of the database, the design of the sprint data table is shown in table 1.

Table 1. Sprint data table

Field name	Data type (length)	Primary key or not	Empty or not
UserID	int	Yes	No
UserName	Varchar(20)	No	No
UserPassword	Varchar(20)	No	Yes
UserType	bit	No	No
StuID	int	Yes	No
StuName	Varchar(20)	No	No
StuNo	Varchar(11)	No	No
StuPassword	Varchar(20)	No	No
TestTypeID	int	Yes	No
TestTypeName	Varchar(20)	No	No
TestTypeDesc	test	No	Yes
TestTypeWeight	int	No	No
TestProjectID	int	Yes	No
TestProjectName	Varchar(20)	No	No
TestProjectTypeID	int	Yes	No

3.3 Decision tree construction algorithm

The construction algorithm of the decision tree can be completed by the training set T , in which it is a training instance when it satisfies the formula (1).

$$T = \{ \langle X, C_j \rangle \}, X = (a_1, a_2, a_3 \dots, a_n) \quad (1)$$

It has n attributes and it is listed in the attribute table (A_1, A_2, \dots, A_n) . a_i represents the value of the attribute A_i . Formula (2) is the classification result of X :

$$C_j \in C = \{ C_1, C_2, \dots, C_m \} \quad (2)$$

The algorithm is divided into the following steps:

Step 1: The attribute A_i is selected as the classification attribute from the attribute table;

Step 2: If the value of the attribute A_i has a total of K_i , then the T is divided into K_i subset T_1, \dots, T_k ;

Step 3: The attribute A_i is deleted from the attribute table. For each T_{ij} , set $T=T_{ij}$;

Step 4: If the attribute table is not empty, then return 1, otherwise output the result.

4 The implementation of the system

The research of this topic uses the C# language in the Visual Studio platform to realize the function development, and the database server selects SQL Server. The realization of the function modules, such as the management of sports performance and the analysis of sports performance, is discussed in detail.

4.1 The realization of the management of sports performance

Sports test category information is the classification management of sports subjects. There will be a certain relationship between it and the sports subjects. Usually, a test type will include a variety of test subjects, and a test subject only belongs to one type of test. The maintenance function of the test type of the system mainly includes the operation of adding, modifying, querying and deleting the basic information and weight information of the sports test type. Taking the deletion function of the sports test type as an example, the function realization is carried out.

The specific implementation process of the deleting function can be described as: The user identifies the type of information that needs to be deleted by checking. When the delete button is clicked, the system responds to the event driver of the button, and the program traverses the type list to get the type number of the user's selection. According to the type number, the sports test subject information table is retrieved to determine whether the current test type includes only the subordinate test subject information. If there is a subordinate subject, the system terminates the current delete operation and pops up the prompt information. If not, the system executes the delete

operation of the test type data. The process of function implementation is shown in figure 4.

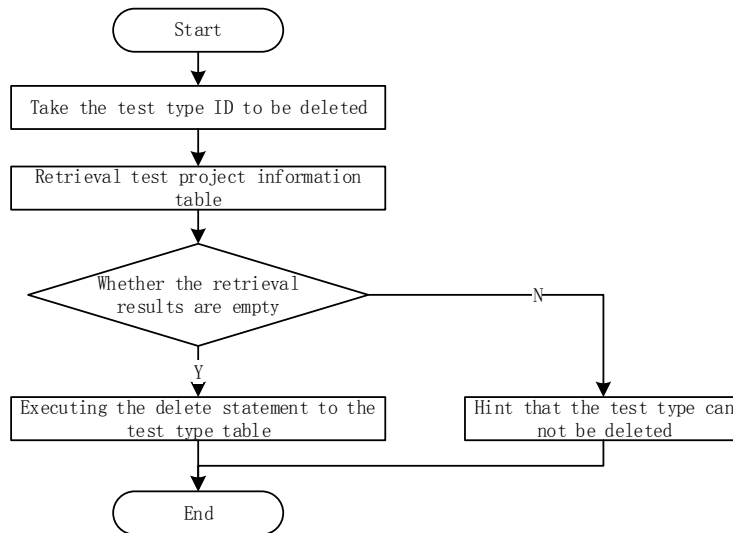


Fig. 4. Test type deleting program flow

The maintenance of sports test project information is based on the type of sports test information. Because the sports test item is not an independent data body, it needs to be attached to the sports test type. Therefore, when carrying out the sports test item maintenance, it is necessary and must choose the sports test type that the current sports test item belongs to. Taking the new function of the sports test project as an example, the realization of the function is introduced. The process of functional implementation is:

The user enters the maintenance interface of the sports test project. In the interface, the user enters the corresponding item information in the corresponding text box or option box according to the hint of the label, and then hits the save button after the input is completed.

In the event driver that saves the button, the user first gets the information in all the text boxes on the page. Then, the user needs to verify the non-empty of required business requirements. The user can proceed to the next step only after ensuring that all required fields have been entered;

When the non-empty validation is passed, the information is filled to the physical entity class. Then, the business operation class is instantiated and the save method is called to perform the storage operation of the data. When the business operation class completes the save operation, it returns the state of the operation, and the page layer determines whether the current operation is successful or not based on the status identity.

The successful management module is a relatively strong information management module in the system, and it is also the core function module in the system. As a matter of fact, performance information is just a string of numbers, while the information associated with it is the core data of the system, such as subject information and student information. The basic operation of the results includes adding, editing, deleting and so on, and its extension function includes many query and statistical operations on the result information. The data generated by these query statistical operations are the important basis for the relevant management departments to make decisions. The following is the introduction of the performance of the record entry and the calculation of the results.

The function of score input is to add students' scores to each sports test subject after the examination of the system's performance data table, and student information must be included in the score information. This examination is the number of sports subject number, and then the specific test score information. The process of function realization is as follows:

First, In the performance entry interface, the administrator should first choose the subjects corresponding to the current entry results, and then choose the students corresponding to the results. The specific fraction information is then entered the performance text box;

Second, after completing the input of the result, the user clicks on the save button to execute the driver of the page event. In the program, the user first gets the information in all the text boxes of the page. Then, the user needs to verify the non-empty of required business requirements. The user can proceed to the next step only after ensuring that the required fields are entered;

Third, When the non-empty validation is passed, the system fills the acquired information into the result entity class. Then, the business operation class is instantiated and the save method is called to perform the storage operation of the data. When the business operation class completes the save operation, it returns the state of the operation, and the page layer determines whether the current operation is successful or not based on the status identity.

4.2 The generation of decision tree

In the process of data mining, the process and result of mining should be expressed by the decision tree. The process of data mining is the process of making the decision tree. Generally, the construction of decision tree mainly consists of two parts: decision tree generation and decision tree pruning. The generation of decision tree is to place data in the root node of the tree, and then to perform recursive operation in a top-down way. The data are scattered to each branch node or leaf node, and finally a complete tree is formed. However, the initial decision tree is not the final result, and the pruning operation needs to be carried out. The pruning of the decision tree is to remove the unrelated data from the decision tree and make the decision tree more scientific and accurate. At this point, it is important to note that the formation of the decision tree should follow two basic principles. The first is that all the data used to construct a decision tree belong to the same category. The second is that when data

segmentation is carried out, the suspension condition is that there is no property to be re-divided.

Recursion is usually used when constructing decision tree. When constructing decision tree, this method only needs to traverse data in database repeatedly. Compared with other methods, constructing decision tree by recursive method has less resources, and it is also suitable for large-scale data mining.

4.3 The application of decision tree in the analysis of students' performance

To explain the application of decision tree in the student sports performance management system, it will be explained through specific cases. Based on the student's sports test result data, the correlation between the students' sprint performance and other sports projects is analyzed.

In the construction of decision tree, first, we should prepare the data and extract the students' sports performance data table. The data table includes basic information and sports performance, such as student ID, name, long jump, long distance running, basketball, sprint, hurdle race, volleyball, shot put and high jump.

After the data is ready, the construction process of the decision tree can be started. The process is as follows:

First, the database table that keeps the student's achievement is normalized. The so-called normalization process is to transform the student performance table into a form which is easy to construct decision tree. In addition, the number "0" and "1" are used to express the students' grades pass or not. During the specific operation, two Excel documents need to be prepared. One document stores the original student data derived from database, and the other document is used to store the basic information of the student. At the same time, the students' achievements are expressed by number "0" and "1". The passing grade is "1" and failure is "0". Through this transformation, the data needed for the decision tree construction is generated.

Second, the training set data is selected from the prepared data. Through the first step, the student's performance is expressed through "pass" or "fail". On this basis, the number of passing students and the number of failed students are counted.

Third, it is necessary to calculate the gain of the information to extract the attributes that are different from the other data in the training set. Because the gain degree directly determines the size of the decision tree, the calculation of the information gain is an important part of the decision tree construction. If the constructed decision tree is too large, the decision ability of the tree will become weaker. On the contrary, the prediction ability of the decision tree will become stronger. The formula for calculating the information gain is as follows:

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \quad (3)$$

Fourth, after calculating the information gain of each sports test project, the construction of the decision tree can be started. When the decision tree is constructed, the first decomposition is started from the top node. The condition of stopping continuing decomposition is that the current node does not have an attribute that can be re-

divided. The information gain mentioned above is the largest. Therefore, long distance race can be regarded as the root node. The properties of long distance race include two kinds: "pass" and "fail". Therefore, two branches can be created under the long-distance running node.

Finally, according to the above methods and principles, some research conclusions can be drawn. If the result of a student's long run fails, the score of the sprint usually fails, and its accuracy is 96.2%. If the student's long distance running passes and the student's hurdle fails, then the student's sprint usually fails. The accuracy rate is 81.3%. If the students' achievements in the long run passes and the hurdles passes, the result of the sprint usually passes, and the accuracy rate is 88%.

The student achievement data and employment data in a class are shown in table 2.

Table 2. Sprint data table

Student	Volleyball	Basketball	Long-distance run	Long jump	Shot-put	Employment situation
A	79	93	67	82	86	A
B	86	74	72	69	73	B
.....
C	75	68	62	71	93	B
D	84	58	73	69	71	D

Using the decision tree construction method introduced in the previous article, the final decision tree is shown in figure 5.

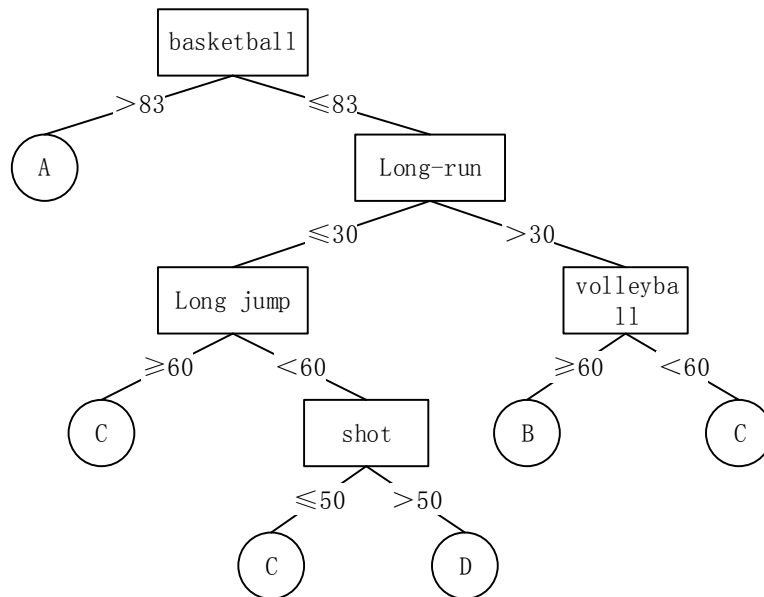


Fig. 5. Generated decision tree

The information provided by the decision tree includes: In all sports test projects, the quality of their employment is determined by the level of basketball results. This shows that the employers are more interested in the basketball hobby of the students when they are employed. In addition, the performance of long distance running has also played a great role in employment. Because the performance of the long run directly reflects the physical health of a student. It shows that the employer will take the students' physique as an important consideration when they employ the employees.

5 Conclusion

The overall structure of the sports performance management system based on decision tree was designed, and a systematic database design plan was proposed. Then, ASP.NET technology was used. Combined with SQL Server database, the following conclusions can be summarized:

First, with the continuous development of information technology in colleges and universities, the use of information technology to serve daily teaching has become an important direction of teaching management reform.

Second, based on ASP.NET technology, SQL Server technology and B/S three tier architecture, a student sports performance management system is designed and implemented.

Third, the overall structure and database structure of the system are introduced in detail. In addition, based on the data mining technology, the data mining for students' sports performance is realized through the decision number algorithm. Taking the correlation of sprint performance, sports performance and the quality of employment as an example, the construction and application process of decision tree are expounded in detail.

Fourth, through the application of data mining, the information of sports performance management is realized, and the efficiency of the performance management is effectively improved. This not only provides a scientific quantitative basis for the improvement of students' physical fitness, but also provides guidance for the improvement of the teaching methods of physical teachers.

6 References

- [1] Hasegawa, T., Terabe, K., Tsuruoka, T., & Aono, M. Atomic switch: atom/ion movement controlled devices for beyond von-neumann computers. *Advanced Materials*, 2012, vol. 24(2), pp. 252-267. <https://doi.org/10.1002/adma.201102597>
- [2] Surhone, L. M., Tennoe, M. T., Henssonow, S. F., Model, R., System, R. D. M., & Codd, E. F., et al. *Relational database*. Cbrg.ethz.ch, 2013, vol. 19(10), pp. 6-32.
- [3] Stuart, D. The data revolution: big data, open data, data infrastructures and their consequences. *Media Culture & Society*, 2016, vol. 37(7), pp. 1110-1111.

- [4] Wang, F., Jing, P., Zhang, S., Ding, F., & Wang, Y. Application of data mining technology in optimizing the extraction process of traditional chinese medicine. *Boletin Tecnico/technical Bulletin*, 2017, vol. 55(10), pp. 522-527.
- [5] Schmitt, J. J., Carranza Leon, D. A., Occhino, J. A., Weaver, A. L., Dowdy, S. C., & Bakum-Gamez, J. N., et al. Determining optimal route of hysterectomy for benign indications: clinical decision tree algorithm. *Obstetrics & Gynecology*, 2017, vol. 129(1), pp. 130. <https://doi.org/10.1097/AOG.0000000000001756>
- [6] Ramanathan, L., Dhanda, S., & Kumar, D. S. Predicting students' performance using modified id3 algorithm. *International Journal of Engineering & Technology*, 2013, vol. 5(3), pp. 2491-2497.
- [7] Niu, X., Shi, F., Hu, X., Xia, J., & Li, N. Predicting the protein solubility by integrating chaos games representation and entropy in information theory. *Expert Systems with Applications An International Journal*, 2014, vol. 41(4), pp. 1672-1679. <https://doi.org/10.1016/j.eswa.2013.08.064>
- [8] Prasad, L. V. N., & Naidu, M. M. Cc-SLIQ: performance enhancement with 2k split points in sliq decision tree algorithm. *Iaeng International Journal of Computer Science*, 2014, vol. 41(3), pp. 163-173.
- [9] Lu, W., & Yan, Z. Application in medical data mining of unmixed interval inductive based SPRINT algorithm. *Journal of Computational Information Systems*, 2014, vol. 10(5), pp. 1811-1819.
- [10] Gao, W. S., Shao, C., & Gao, Q. Pseudo-collision in swarm optimization algorithm and solution: rain forest algorithm. *Acta Physica Sinica*, 2013, vol. 62(19), pp. 750-754.

7 Author

Jian Chen is with the Hunan Vocational College of Science and Technology, Hunan, China.

Article submitted 27 February 2018. Resubmitted 03 April 2018. Final acceptance 05 June 2018. Final version published as submitted by the author.