

Design of an Instant Data Analysis System for Sports Training Based on Data Mining Technology

QunBi Lei, Hunan Vocational and Technical College of Environmental Biology, China*

ABSTRACT

Data mining (DM) is an in-depth approach to data analysis by mining useful information from large amounts of data, and this technique is now being used in an increasing number of fields. In this paper, the authors present the design of a real-time data analysis system for sports training based on DM technology and use the corresponding mining tools of DM technology to discover relevant patterns or laws hidden in the data. Therefore, using the real-time data analysis system for sports training based on DM technology, useful information and patterns for improving examination performance can be obtained, which can improve targeted teaching methods and help students overcome learning difficulties, providing rational teaching, synchronizing courses, establishing preparation, effectively guiding students in course selection, and improving course quality and educational effectiveness.

KEYWORDS

DM, Physical Training, Real-Time Data Analysis System

INTRODUCTION

The quality of education offered by a school directly affects the vitality of that school, whose core mission is to educate its students (Hui & Jin, 2021). Sport is a highly integrated discipline, encompassing athletics, humanities, and social sciences (Yin & Cui, 2021). In the past, the research field of real-time data analysis systems was limited to exercise training, exercise evaluation, and exercise management. (Rajšp & Fister, 2020). In recent years, guided by the principle that science and technology are primary drivers of productivity, China's sports departments have increasingly recognized the importance of technological advancement, leading to enhanced scientific decision making and management in sports (Bonidia et al., 2018). Moreover, as the scale of education expands, the training process generates vast quantities of data, making this data increasingly difficult for educational decision makers to understand. Traditional data processing methods are ill-equipped to handle this surge in data accumulation (Karachi et al., 2017). Given this state of affairs, teachers must convert scores recorded in minutes, seconds, and meters to percentages based on national physical

DOI: 10.4018/IJWLTT.330991

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

fitness standards (Afzali & Mohammadi, 2018). At the same time, each teacher must record the converted scores using the classroom's educational management software (Zhang & Mao, 2021). With these vast amounts of data, the existing database management methods and data statistics methods are increasingly unable to adapt to the national proposal of "healthy exercise" and sports talent stratification (Hu, 2018). Data based on various sports indicators lacks interconnectedness (Wang & Chen, 2017). Therefore, if we do not adopt advanced management concepts, change the school philosophy, deeply understand the diverse needs of society, understand the characteristics of each student, and incorporate specific specialties, it will be difficult to apply the original management methods and teaching methods, and sustainable development will become increasingly challenging (Gao et al., 2018). Current methods offer only basic queries and statistics, without in-depth analysis to identify factors impacting student performance (Wang & Liang, 2021). Data mining (DM) and data warehousing techniques can better achieve this latter objective (Gamonales et al., 2021). DM excels in both identifying and solving problems. It is the process of extracting hidden but potentially useful information and knowledge from large amounts of noisy real-world application data (Choi & Yoon, 2017). DM and knowledge discovery have advanced data processing techniques (Kantilal & Sharma, 2020). It not only allows in-depth analysis of data, but also provides the necessary information in a timely and accurate manner, which enables the deep search for interrelationships between various elements within an immense amount of data (Park et al., 2020), as well as guiding the creation of new rules for learning and classroom training in school sports. A large amount of sports data is being accumulated in the fields of sports competition and sports industry. Sports researchers now face the vital task of using these data to discover information that is useful but easily overlooked.

The innovations of this paper are:

- (1) The aim of researching a real-time data analysis system for PE based on DM technology is to boost teachers' efficiency and accuracy, thus freeing them from monotonous tasks.
- (2) Using DM helps identify underlying factors that impact educators' teaching, thus providing suggestions for improving the quality of teaching.
- (3) DM technology is applied to a real-time data analysis system in the field of physical training, using a large amount of experimental data to establish a data warehouse that matches the physical capabilities of college students.

The paper is structured as follows:

The below section presents a summary of the research related to DM technology, a real-time data analysis system for sports training. Next, the paper details the design of the real-time data analysis system for DM, describing its functional modules for clarity. The paper then delves into the application of DM technology in the sports training instant data analysis system, focusing on two aspects: data input of DM, and DM analysis in the instant data analysis system for sports training. Lastly, a conclusion is provided regarding the findings of the study.

RELATED WORKS

Sports Real-Time Data Analysis System

A survey of relevant data shows that the physical condition of Chinese youth has been declining continuously. Exploring the application of data warehousing and DM technology in practical courses, for the purposes of information management and a deeper analysis of student sports data, is a vital area of study for sports researchers to enhance their understanding. Several scholars have delved into real-time sports data analysis systems, national fitness information assessments, and the sophistication of such systems. This growing interest underscores the recognition of DM's robust capabilities and the desire to harness them.

Lili (2018) proposed an instant data analysis system component structure, including dialogue, data (database and its management), and model-based components. Pan (2019) introduced the ID3 algorithm to extract and analyze data associated with school administration. By deeply exploring mining techniques, he identified relationships between curricula, offering valuable insights for school decision making. Rajput and Thakur (2019) introduced the framework of Decision Support Systems (DSS), encompassing the linguistic system, problem processing system, and knowledge system. These components possess distinct characteristics and exert influence to a certain degree. Tsai used DM techniques to analyze a large amount of accumulated historical data. Factors affecting excavation, school sport research, and education training were identified, and the relationships between these factors were used to identify sport talents (Li et al., 2021). Patel and Shah (2018) suggested integrating the instant data analysis system with an expert system, enhancing the capabilities of the DSS to more comprehensively represent the decision-making process. This approach effectively addresses semi-structured and unstructured challenges. To initiate this, they developed an intelligent real-time data analysis system.

To holistically address the needs of college students, scientific training theories and advanced methods can be integrated into college sports training management. Furthermore, applying DM technology to support instant data analysis in college sports training is theoretically feasible.

DATA MINING

DM is a groundbreaking technology, bridging various disciplines and enhancing decision support. In order to organically integrate the experience of college students, advanced training methods rooted in scientific training theories can be implemented in collegiate sports training management. DM mainly integrates database technology, artificial intelligence, machine learning, statistics, visualization techniques, pattern recognition, and other information technologies, in order to analyze data in a highly automated manner, perform inductive reasoning, and discover potential patterns, thereby allowing decision makers to adjust strategies, reduce risks, and make informed decisions.

Qin and Min (2020) used various DM methods such as classification, clustering, and sequential pattern analysis to mine information from online education databases and explore the relationship between student performance and learning behavior, thus helping to improve the quality of online education. Using the DM technology provided by IBM, L. Xu et al. (2017) improved users' application of data, which progressed from low-level end-query operations to the provision of beneficial decision support for business decision makers at all levels. Y. Xu et al. (2017) used DM technology to implement personalized education in a distance learning system so that the resources of this system could be configured to support the personalized learning needs of students. Bandaru et al. (2017) investigated the application of principal component analysis techniques and the Bayesian nearest neighbor algorithm in DM and used these tools in the evaluation of graduate students' grades, to reduce the number of indicators and the amount information requiring analysis. Slater et al. (2017) studied faculty information databases and used association rule mining techniques to find relationships between various factors that influence academic development.

Educational DM transforms raw data from diverse educational systems into actionable insights. Applying DM to the sports training instant data analysis system promises compelling outcomes.

DESIGN IDEAS FOR REAL-TIME DATA ANALYSIS SYSTEM FOR PHYSICAL TRAINING BASED ON DM TECHNOLOGY

Design of Real-Time Data Analysis System Based on DM

Schools, both domestically and internationally, along with different societal sectors, have established a series of online information services, which can also be called instant data analysis systems (Jassim

& Abdulwahid, 2021). Such a system aims to enhance the efficiency and accuracy of PE teachers in higher education, relieving them from repetitive tasks. Consequently, it enhances their teaching quality and management skills and meets the needs of three types of users: administrators, teachers, and students, by providing a more accurate and comprehensive information platform (Praveena et al., 2019). SPSS, a statistical software, integrates data collation, analysis, and result output, using an intuitive interface, and the output can be tabulated graphically. The system allows the registration of information, inquiries, and online question and answer requirements, thus enabling internal personnel to learn various information related to the school or enterprise without having to leave home. Also, when comparing and analyzing factor data of different dimensions, it is necessary to standardize them. The values of i and j factors in all samples are:

$$Z_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j} \quad (1)$$

μ_j — Mean value of i th factor;
 σ_j — The standard deviation of j th factor

First of all, the system adopts the B/S (Browser/Server) model, which is a network structure model developed after the rise of the web, and is an improved variation of the C/S model. Different user roles have different operation rights in the system. The model provides a convenient and user-friendly interactive environment for decision makers to input the data on student physical quality indexes for a class or major, according to the prompts of the interface, and then submit this data to the server for processing. In the later stage of the algorithm search, the discovery probability is reduced in order to increase the convergence speed of the algorithm. Therefore, the discovery probability is improved as:

$$pa_t = \exp(t / t_{\max}) * \cos(t / t_{\max}) * pa_{begin} \quad (2)$$

pa_t — Discovery probability of t iteration
 $\exp(t / t_{\max}) * \cos(t / t_{\max})$ — Function dynamic decreasing factor

Therefore, a mathematical model is required to describe the different network connections from a theoretical standpoint. Additionally, to facilitate analysis in high-level languages, appropriate data structures must be identified (Yang, 2021). This method is ineffective when used to find classes that differ greatly in size, or classes that present a non-convex shape, and it is sensitive to noise and outlier data. Judging the probability of categorizing an instance into a class based on the value of the independent variable:

$$P(X) = P(Y = 1 | X) = 1 / (1 + e^{f(x)}) \quad (3)$$

X — Collection of data description attributes type
 Y — Represents the category attribute of data
 $P(X)$ — Conditional probability

Second, a web browser like Internet Explorer serves as the primary application software, and only one browser, such as Internet Explorer, need be installed on the client machine as a unified client. The core part of system function implementation is centralized on the server, thus simplifying the development, maintenance, and use of the system. Before proceeding, one must check if specific test items exist within this test category. If they do, they cannot be deleted, otherwise the deletion operation is executed. Figure 1 illustrates the program flow for implementing this function.

For various test types, each type contributes a specific percentage to the total score. The system should offer functionalities to add, edit, or delete these test types and adjust the weighting of the score for each type (Dai & Li, 2021). All tuning methods have limited effect on improving database performance if the database logic is poorly designed. Figure 2 illustrates the flow of DM.

The system provides users with insights into its operational status, execution results, and reasoning outcomes. When describing the pattern of linking relationships between pages, a matrix representation can be used. Since its inception, DM technology has always been application-oriented. The error is propagated in the reverse direction by constantly updating the weights and the bias of the prediction error of the representation network. The error is calculated as:

Figure 1. Process of test type deletion program

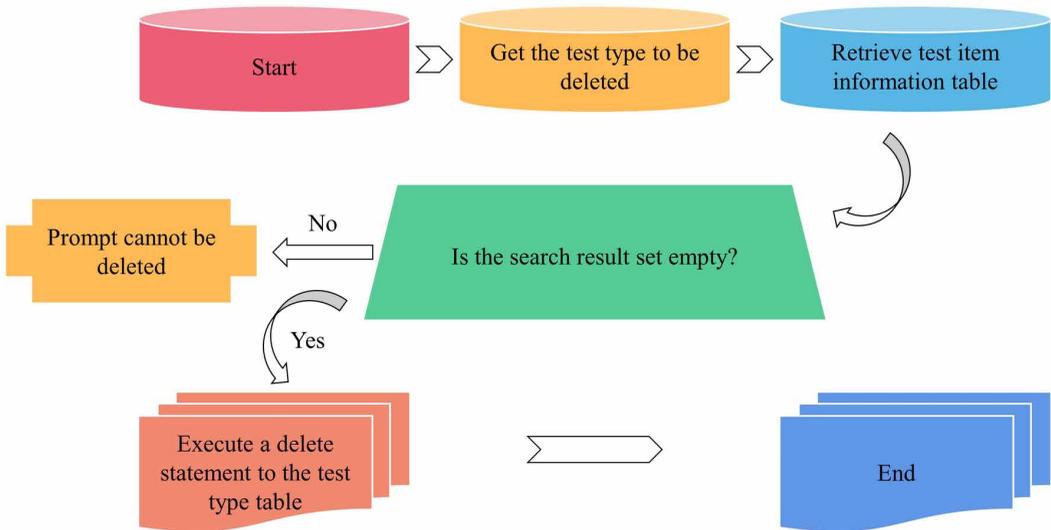
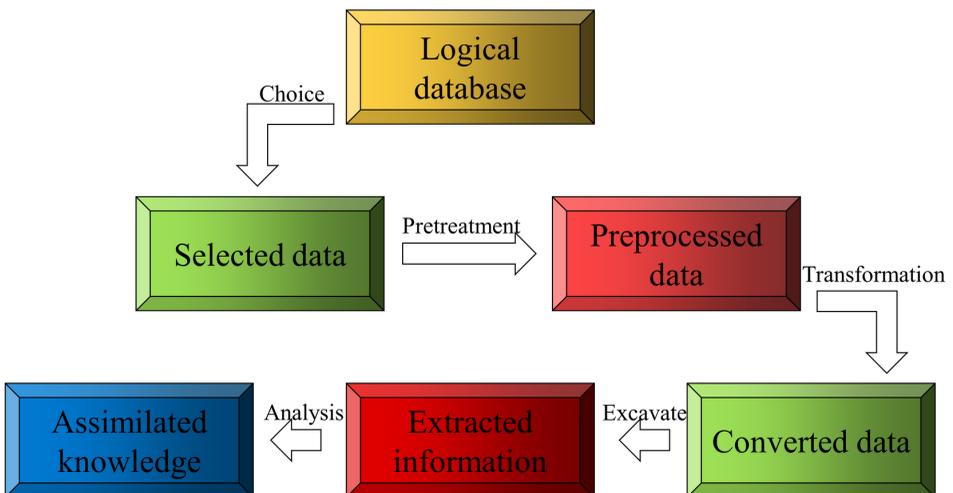


Figure 2. The process of DM



$$Err_j = O_j(1 - O_j)(T_j - O_j) \quad (4)$$

j — Output layer node

E_{errj} — Error

O_j — Actual output

T_j — Based on the known target value of the given training sample

The system adopts a three-tier B/S architecture. On the client side, users access via a browser, while the server side features a three-tiered design: representation layer, service application layer, and data storage layer. Users, when logging into the system, must provide credentials like a username and password. Only after verification can they access their respective system backend for related operations. In the case of merging two sets, it is only necessary to perform a bitwise summation operation on the binary sequences representing the two sets. The binary sequences are used to perform the operation as follows:

$$N(h) = \frac{1}{0.77351} \sum_{i \in V} 2^{\frac{\sum_{i=1}^k b_i(i)}{k}} \quad (5)$$

i — Node

b_i — FMbitmask

For each test type, new items can be added, and existing ones can be modified or deleted. The corresponding training programs are generated according to the user's requirements, including individual training suggestions and group training programs. According to the central object of each cluster, i.e., the mean value of data records of that class, the size of each data record from those central objects is calculated; then the corresponding data records are reclassified according to the minimum distance.

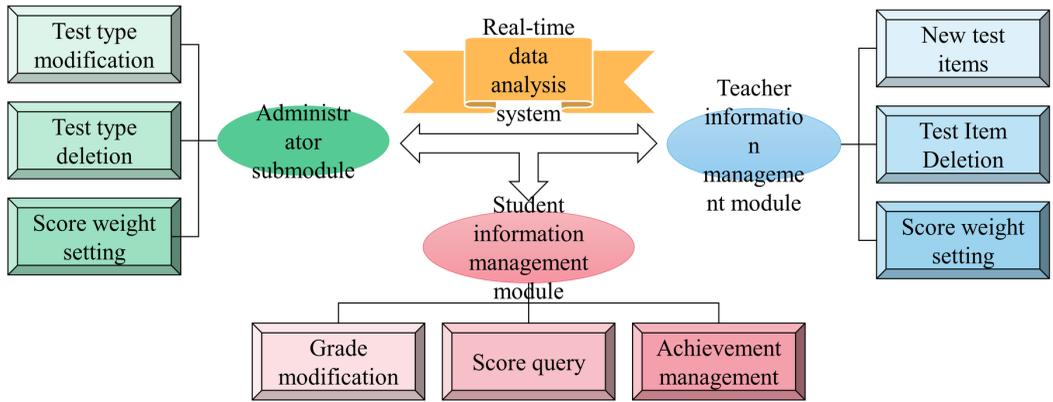
Design of Functional Module of Real-Time Data Analysis System

In the DM query language, five basic DM primitives are defined: task-related data primitives, primitives for type of knowledge being mined, background knowledge primitives, interest degree measurement primitives, representation of discovered patterns, and visualization primitives (Liang, 2021). Within this system, the three main categories of users targeted are administrators, teachers, and students; therefore, for each category of users, their respective instant data analysis systems are created separately. The main functions included in each subsystem are described below. The system functional module diagram is shown in Figure 3.

Firstly, the administrator sub-module mainly includes permission management, teacher information management, student information management, and grade analysis. Sub-functional modules can be treated as individual units, each with their own production cycle; the biggest advantage of this approach is the division of labor in software development. For example, in the target system, the student grade management module can be divided into sub-functions such as student grade entry, grade modification, grade deletion, and grade query, and different functions can be assigned to individual developers for more focused development. The mean substitution method is utilized to determine the setup and extract the load data:

$$\bar{A}(d, t) = \frac{1}{n} \sum_{d=1}^n A(d, t) \quad (6)$$

Figure 3. System function module diagram



$\bar{A}(d, t)$ — Average value of load

All necessary information for decision making is systematically provided. However, for mid-level and senior decision makers, only a fraction of the required internal and external news, experiences, and judgments tailored to individual decision-making styles are available. This makes it challenging for them to operate comfortably. As a result, data such as sports performance metrics and physical fitness questionnaires from students had to be gathered and inputted. This also indicates that the solution to the linear programming problem yields an integer solution. By converting the diversity constraint into an upper and lower bound constraint for classification, the optimization problem can be defined as follows:

$$\max \sum_{i,j} w_{ij} x_{ij} \quad (7)$$

w_{ij} — ij th preference scoring matrix

x_{ij} — Final match.

Following this is the teacher information management module, which mainly includes information such as basic teacher information, course information for the teacher's class, information on the students taught by the teacher, and the teacher's analysis of student performance.

The sub-functional modules can be circulated as separate products in the market; for example, in this system, the analysis of sports results can be completely independent of the sports results instant data analysis system itself. The collected source data is then subjected to a series of cleanups and transformations to remove invalid and erroneous data. The deviation between predicted and actual values is measured using the root mean square error, which reflects the dispersion degree of a dataset:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\sigma_i - \hat{\sigma}_i)^2}{n}}, i = 1, 2, \dots, n \quad (8)$$

$\hat{\sigma}_i$ — A predicted value in the set of predicted results

σ_i — The corresponding actual load value

n — The number of predicted values in the prediction result set

To conserve memory, we forgo the recursive method in favor of a hierarchical reconstruction. Using a LinkedList, we construct the FPTREE layer by layer. When queuing related nodes, we retain only essential association information, releasing other memory-intensive details, thus further decreasing memory usage.

The student information management module primarily includes admission results, class details, and academic performance. Optionally, records of school rewards, punishments, extracurricular activities, educational experiences, graduation data, and employment information can be added. Feedback from alumni regarding their academic experiences might also be incorporated. Each sub-functional module can operate independently. With extensive raw data, the division method can be employed, allowing the FP-Tree to fit into the primary memory. The level-weighted support of the item set $X = \{i_1, i_2, \dots, i_k\}$ is defined as:

$$Sup_h = \max\{h_1 \cdot h_2 \cdot \dots \cdot h_k\} \times Sup(X) \quad (9)$$

$Sup(X)$ — Traditional support count of item X ;
 $Max\{h_1, h_2, \dots, h_k\}$ — Weight of items

Since the different functional subsystems can be logically independent, independent maintenance can be achieved. Then a multidimensional dataset based on this data source is established in SQLServer to facilitate the maintenance and analysis of the life cycle of the network service. For each period of the service, the system has to estimate the BER of the network service and retain the BER detection value. By assessing correlations among N-gram features of varying granularities, we assign weights to these features. Their representations are derived through a weighted summation:

$$\alpha_i^t = \frac{\exp(U_i^{tT} u_w)}{\sum_t \exp(U_i^{tT} u_w)} \quad (10)$$

c_i^t — The i word representation of t feature

Therefore, the scattered data objects in the class are selected and later shrunk to the center of the class according to the shrinkage factor (which usually takes the value of a specific fraction). This provides DM systems with a law to follow in terms of model definition and description. Various DM systems can share the model, and DM models can be nested in the middle of application systems. It is possible for DM to accomplish deep mining without separate development.

ANALYZING THE APPLICATION OF DM TECHNOLOGY IN THE REAL-TIME DATA ANALYSIS SYSTEM FOR PHYSICAL TRAINING

Data Import DM Algorithm Processing and Analysis

The data required for DM analysis is imported into this system from the designed and organized data warehouse for the corresponding DM operations, and the specific algorithms are processed as follows.

Initially, the data is represented using a matrix approach. A graph, influenced by the decision of its proximate active associations, can be depicted as a matrix. The network matrix can be formed by arranging each node in rows and columns. The matrix is significant for the mining of attributes such as service name, rate, path, and current BER. When analyzing the temporality of BER of optical channels, the matrix can provide reference for the assurance of service quality and decision making of managers. Students can compare and analyze their academic performance with peers in the same class or other classes and majors. This insight helps pinpoint areas of potential improvement, enabling the development of more effective study strategies. We conducted random sampling using predetermined

seeds. Valid data was split into training and test samples at an 80%–20% ratio. Figures 4 and 5 display the information gain and information gain rate of each attribute, respectively.

For any new object, the COBWEB algorithm traces a suitable path downward, adjusting the count to determine the optimal node for classification. Rather than employing the average value of a cluster as the new centroid, the object most central to the cluster is selected. The basic strategy is to first select a representative object for each cluster and divide the other objects into different clusters according to the division criterion.

As a secondary step, all entities are grouped into one cluster. This is then progressively split into smaller clusters until each entity has its cluster or until a specific user-defined condition is met, such as reaching a set number of clusters or when the distance between the nearest clusters surpasses a set threshold. The evaluation criterion for finding the best node is to temporarily place the object at each node position and then calculate the magnitude of the resulting classification utility. The position with the maximum classification utility is a good choice for the object. A specific BER value is chosen for estimating the optical channel, and this BER might recur, as illustrated in Table 1.

Initially, data is segmented into subsets. We then employ the decision tree algorithm to traverse these subsets, constructing a hierarchical decision tree from the top down. After the tree—with its roots, nodes, and leaves—is constructed, nodes and leaves undergo pruning based on the algorithm. Following this, representative points are swapped with non-representative ones; if the clustering quality improves, the swap is retained. This approach assists athletes in identifying and addressing weak points in their training. By focusing on strengths during competition and avoiding pitfalls, athletes can better control the situation and optimize performance.

Crucially, for effective clustering, it's essential to distinguish primary data, which plays a pivotal role, from secondary data. Students can periodically evaluate and provide feedback on teaching methods, assisting teachers in refining their approach. Additionally, this allows students to analyze their performance and offer suggestions. The system autonomously extracts valuable insights from the data, formulating decision rules via knowledge reduction and dependency analysis. Initially, every object is treated as its cluster; as similarity assessments progress, larger unified clusters form.

Figure 4. Information gain of each attribute

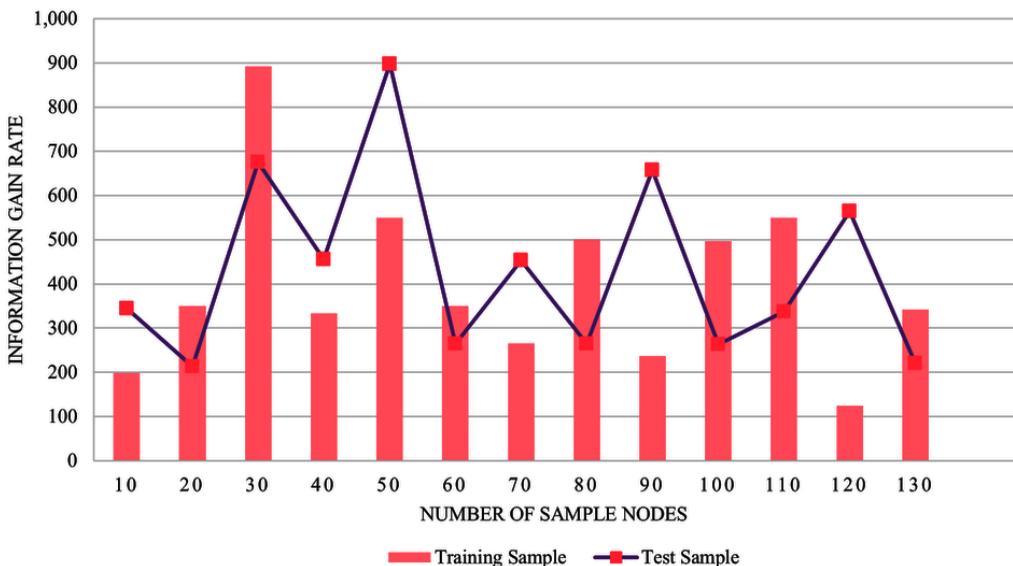


Figure 5. Information gain rate of each attribute

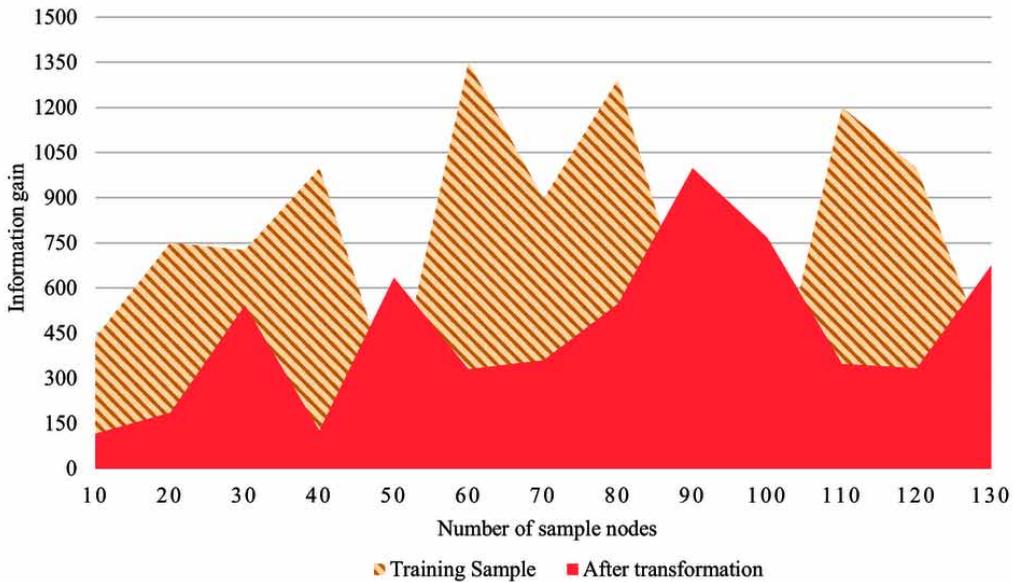


Table 1. Estimated data error rate and measured data error rate

Business Number	1520	1521	1523	1524	1525	1526
Estimated error rate	1.60e-5	1.60e-7	1.60e-3	1.60e-3	1.60e-9	1.60e-1
Measured error rate	1.60e-2	1.60e-6	1.60e-7	1.60e-4	1.60e-6	1.60e-2

Given that the database stores data objects such as transactions and text, it might be categorized as a transactional database, text database, or object-oriented database, among others.

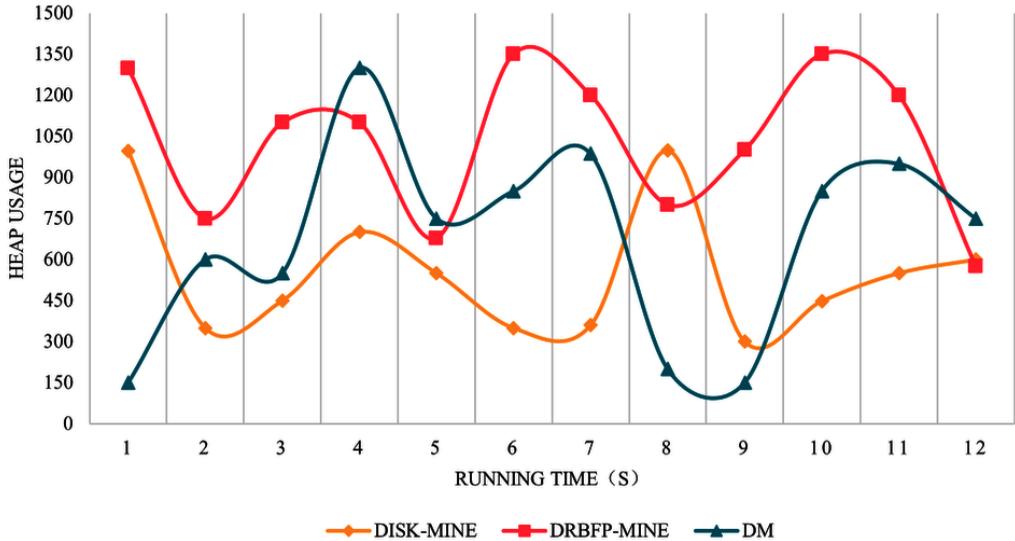
DM Analysis in Training Real-Time Data Analysis System

Embracing the modularity principle requires a focus on information encapsulation and localization. Large datasets can harbor numerous correlation rules. Indiscriminate mining can yield inefficiencies and irrelevant rules. Hence, strategic association rule mining is paramount in real-time data analysis systems.

Employing constraints enables targeted association rule mining. Relational databases, accessible via structured SQL statements, facilitate operations like joining, selection, and projection. Advanced operations, underpinned by sound mathematical and statistical foundations, further underscore the prevalence of relational databases in DM data organization. The efficiencies of extensive data algorithms like DISK-MINE, DRBFP-MINE, and DM are illustrated by bar records generated from 30,000 unique entries (refer to Figure 6).

Depending on specific application domain characteristics, various algorithms may be used for the same problem. Descriptive mining tasks, for instance, can be tackled with conceptual abstractions, including data characterization, differentiation, and summarization. Real-time scoring data, a type of structured big data, hints at potential computational challenges with SQL queries—underscoring the need for innovative querying methods.

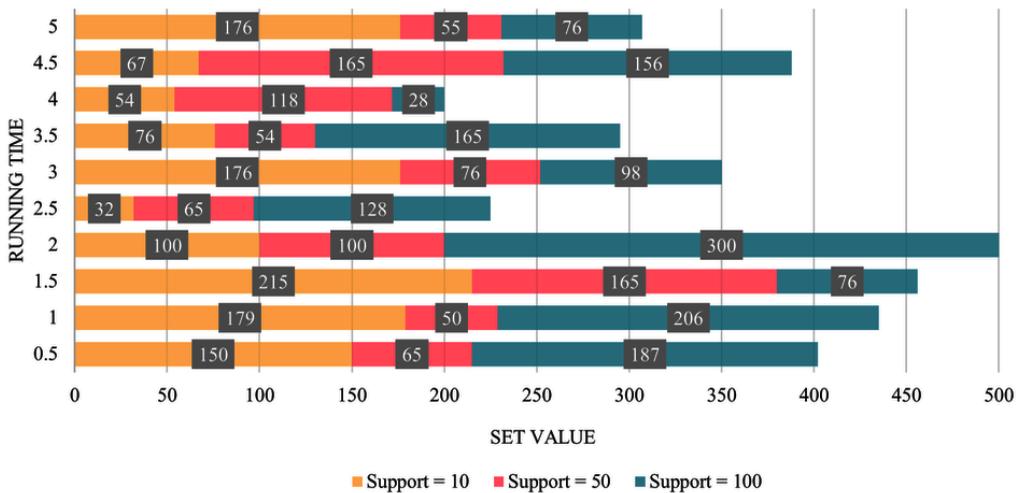
Figure 6. Comparative analysis of performance of large-scale dataset algorithms



Furthermore, frequent set queries employ constraints. These constraints, pivotal in the mining process, can be grouped as monotonicity, anti-monotonicity, transformable, and conciseness constraints. This method, offering an intuitive visualization of data rules, boasts enhanced accuracy, efficiency, and user comprehension. Data structures often comprise multiple resolution levels, forming a hierarchical grid. When the support degree is set at 10, 50, or 100, and each level is operated 15 times, the average execution time for the statistical record is depicted in Figure 7.

Regarding data access, applications should ensure streamlined access to datasets via HDFS. Designed to prioritize data throughput, HDFS offers superior performance by somewhat relaxing POSIX compliance.

Figure 7. Running time comparison



Ultimately, merging temporal intervals consolidates potential post-filtering fragments into distinct mining time zones. Each zone then undergoes memory-based mining to produce association rules. The mining outcomes for the dataset 'Dataset 1' using DM and GF-DBSCAN are detailed in Table 2.

From the data presented in the table above, DM's average run-time for the dataset 'Dataset 1' shows an improvement of 0.081 compared to GF-DBSCAN. This underscores DM's growing efficiency regarding time complexity.

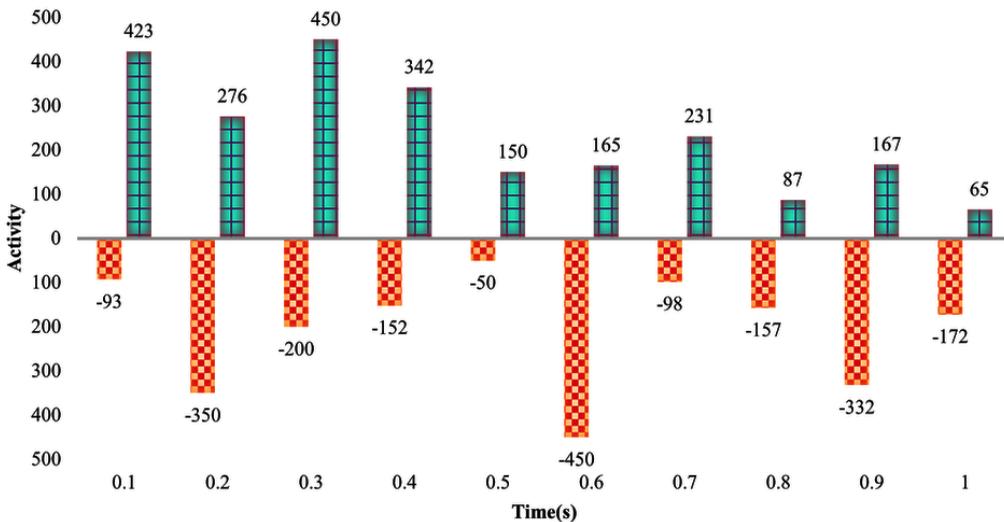
For intricate DM tasks, which entail multiple data sources and DM modules, the exchange of results between these modules is essential. PMML (Predictive Model Markup Language) serves this need effectively. Its primary components offer a flexible model exchange and data format conversion, ensuring the model remains distinct from the associated data and tools. Subsequent objects are grouped based on their likeness to cluster centroids. The mean value within each cluster is then used as the new centroid. The computation is reiterated until the criterion function converges. To showcase DM's efficiency, the program's maximum size is configured to mine association rules from expansive datasets within a minimized memory allocation. The comparison between the system's CPU and GC activities is illustrated in Figure 8.

The unique information within each cell allows data extraction without depending on aggregated query information, offering a computation process that remains query-independent. The grid structure also facilitates parallel processing and incremental updates. The primary objectives of a data warehouse encompass aiding users with the administration of the data warehouse server and catering to their query requirements through its specialized storage architecture. The underlying algorithm, rooted in statistics, exhibits optimal performance and robustness in the case of datasets smaller than a few

Table 2. Running time of DM and GF-DBSCAN on Dataset 1

Running Time(s)	2000	4000	6000
DM	0.189	0.245	0.318
GF-DBSCAN	0.231	0.337	0.429

Figure 8. Comparison of CPU and GC activities



hundred. Moreover, the system can be dynamically scaled to match application demands without necessitating interruptions, thus streamlining configuration and management.

CONCLUSION

Physical Education (PE) is a crucial component of academic curricula. The integration of computer and network technologies into daily PE routines has become indispensable in our digitally driven era. DM Instant Data Analysis Systems elevate research and development, offering a competitive edge to coaches and athletes. Evaluating the need for data processing and analysis in undergraduate curricula, we can gauge the real-world applicability of courses and their contribution to professional competence during university education. DM has gained prominence not only in pattern recognition and machine learning but also stands as a lucrative research area for many professionals. The proposed real-time data analysis system for PE, rooted in DM technology, seeks to enhance educators' methodologies and overall instructional quality. Simultaneously, it aims to foster students' learning habits and efficiency. By equipping educators with actionable insights, students are better positioned to identify their learning gaps, thus nurturing a self-directed learning environment.

DATA AVAILABILITY

The figures and table used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The author declare that he has no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGMENT

The author would like to show sincere thanks to those techniques who have contributed to this research.

AUTHOR NOTE

The author is sincerely grateful to the developers of those methods that have contributed to this research.

The author declares there is no conflict of interest.

This work was not supported by any grant funding.

REFERENCES

- Afzali, G. A., & Mohammadi, S. (2018). Privacy preserving big data mining: Association rule hiding using fuzzy logic approach. *IET Information Security*, 12(1), 15–24. doi:10.1049/iet-ifs.2015.0545
- Bandaru, S., Ng, A. H., & Deb, K. (2017). Data mining methods for knowledge discovery in multi-objective optimization: Part A-Survey. *Expert Systems with Applications*, 70, 139–159. doi:10.1016/j.eswa.2016.10.015
- Bonidia, R. P., Brancher, J. D., & Busto, R. M. (2018). Data mining in sports: A systematic review. *Revista IEEE América Latina*, 16(1), 232–239. doi:10.1109/TLA.2018.8291478
- Choi, C. H., & Yoon, G. Y. (2017). Cycling winner prediction model by using match information: Application of decision tree analysis based on data mining. *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, 19(4), 15–26. doi:10.21797/ksme.2017.19.4.002
- Dai, X., & Li, S. (2021). Volleyball data analysis system and method based on machine learning. *Wireless Communications and Mobile Computing*, 2021(20), 1–11. doi:10.1155/2021/9943067
- Gamonales, J. M., León, K., Rojas-Valverde, D., Sánchez-Ureña, B., & Muñoz-Jiménez, J. (2021). Data mining to select relevant variables influencing external and internal workload of elite blind 5-a-side soccer. *International Journal of Environmental Research and Public Health*, 18(6), 3155. doi:10.3390/ijerph18063155 PMID:33803780
- Gao, Q., Zhang, F. L., & Wang, R. J. (2018). Mining frequent sets using fuzzy multiple-level association rules. *Journal of Electronic Science and Technology*, 16(2), 145–152.
- Hu, Q. (2018). Research on the development of tennis sports based on data mining. *IPPTA*, 30(8), 179–182.
- Hui, D., & Jin, W. (2021). Design of real-time data analysis system for physical training based on data mining technology. *Journal of Physics*, 1982(1), 012206.
- Jassim, M. A., & Abdulwahid, S. N. (2021). Data mining preparation: Process, techniques and major issues in data analysis. *IOP Conference Series. Materials Science and Engineering*, 1090(1), 012053. doi:10.1088/1757-899X/1090/1/012053
- Kantilal, S. Y., & Sharma, Y. K. (2020). Understanding association rule in data mining. *International Journal of Advanced Research*, 8(6), 289–292. doi:10.21474/IJAR01/11097
- Karachi, A., Dezfuli, M. G., & Haghjoo, M. S. (2017). Intelligent information and database systems. *Lecture Notes in Computer Science*, 5990(6), 891–896.
- Li, J., Lei, H., & Tsai, S. B. (2021). Online data migration model and ID3 algorithm in sports competition action data mining application. *Wireless Communications and Mobile Computing*, 2021(7), 1–11.
- Liang, M. (2021). Optimization of quantitative financial data analysis system based on deep learning. *Complexity*, 2021(1), 1–11. doi:10.1155/2021/5527615
- Lili, X. (2018). Redundant processing algorithm of association rules based on hyper graph in data mining. *IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association*, 30(6), 683–689.
- Pan, L. (2019). A big data-based data mining tool for physical education and technical and tactical analysis. *International Journal of Emerging Technologies in Learning*, 14(22), 220–231. doi:10.3991/ijet.v14i22.11345
- Park, S. U., Ahn, H., Kim, D. K., & So, W. Y. (2020). Big data analysis of sports and physical activities among Korean adolescents. *International Journal of Environmental Research and Public Health*, 17(15), 5577. doi:10.3390/ijerph17155577 PMID:32748858
- Patel, J., & Shah, P. (2018). A review approaches for hiding sensitive association rules in data mining. *International Journal on Computer Science and Engineering*, 6(11), 920–924.
- Praveena, K., Sirisha, G., Babu, S. S., & Rao, P. S. (2019). Efficient method in association rule hiding for privacy preserving with data mining approach. *Ingénierie des Systèmes d'Information*, 24(1), 47–50. doi:10.18280/isi.240106
- Qin, Y., & Min, G. (2020). Outlier data mining of multivariate time series based on association rule mapping. *International Journal of Internet Manufacturing and Services*, 7(1–2), 83–96. doi:10.1504/IJIMS.2020.105038

- Rajput, V., & Thakur, P. A. (2019). Improving data analysis using data mining techniques for KSOMM and PAPLM. *International Journal of Advanced Research in Computer and Communication Engineering*, 8(6), 28–32. doi:10.17148/IJARCCCE.2019.8606
- Rajšp, A., & Fister, I. Jr. (2020). A systematic literature review of intelligent data analysis methods for smart sport training. *Applied Sciences (Basel, Switzerland)*, 10(9), 3013. doi:10.3390/app10093013
- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, 42(1), 85–106. doi:10.3102/1076998616666808
- Wang, D., & Liang, F. (2021). Application of artificial intelligence and big data in sports event service—Take Guilin as an example. *Journal of Physics: Conference Series*, 1881(3), 032056. doi:10.1088/1742-6596/1881/3/032056
- Wang, Z., & Chen, D. (2017). An optimized data mining algorithm application in volleyball match technique and competition tactics analysis. *Boletin Tecnico/Technical Bulletin*, 55(17), 782–787.
- Xu, L., Jiang, C., & Wang, J. (2017a). Information security in big data: Privacy and data mining. *IEEE Access : Practical Innovations, Open Solutions*, 2(2), 1149–1176.
- Xu, Y., Wu, D., & Xue, J. (2017). Research on the framework of data processing and sustainable development of sports events based on data mining technology. *Boletin Tecnico/Technical Bulletin*, 55(18), 338–344.
- Yang, N. (2021). Discussion on the application of enterprise business intelligence data analysis system. *Journal of Electronic Research and Application*, 5(4), 23–26. doi:10.26689/jera.v5i4.2520
- Yin, Z., & Cui, W. (2021). Outlier data mining model for sports data analysis. *Journal of Intelligent & Fuzzy Systems*, 40(2), 2733–2742. doi:10.3233/JIFS-189315
- Zhang, S., & Mao, H. (2021). Optimization analysis of tennis players' physical fitness index based on data mining and mobile computing. *Wireless Communications and Mobile Computing*, 2021(11), 1–11. doi:10.1155/2021/9838477