



# Design and realization of data mining simulation and methodological models

Zhang Zong Biao

*Institute of Modern Services, Zhejiang Shuren University, China*

## ARTICLE INFO

### Keywords:

Data mining  
Methodological models  
Statistical Models  
Python

## ABSTRACT

Data mining in education is steadily gaining momentum to explore the learning analytics because of its easy to use with any programming knowledge. This emerging method primarily focus on developing methods for handling and exploring voluminous information represented in distinct types of data. Deployment of data mining in education helps to better understand the synergy between the learning and educational settings established by the institutions to foster their learning activity. This work presents a data mining based virtual simulation experimental teaching system which provisions personalized interactive learning, real-world scene experimental training and comprehensive assessment of the learning effect. The proposed system employs interactive programming, automatic evaluation of experimental results, assessing the impact of teaching effectiveness with other modules. The work also provides a series of comprehensive cases and programming experiments apart from presenting the execution pages. The results indicate that this system can intuitively help students to improve their learning process, student participation and foster independent learning ability. The application of data mining alleviates the dependency on complex programming paradigms to make learning analytics, which is beneficial to both the teaching and learning communities.

## 1. Introduction

With the development of artificial intelligence and big data technology, the demand for big data professionals in various fields has shown a huge growth trend. Governments, organizations, and enterprises are all committed to mining the huge value behind data, finding correlations and pattern analysis, predicting future trends, visualizing data, and assisting refined management and intelligent decision-making Zheng (Ageed et al., 2021). However, as an emerging technology, the lack of a corresponding experimental environment and platform has led to a serious disconnect between theory and practice. How to effectively allow students to conduct practical experiments, deepen their understanding of abstract data mining knowledge points, and master data mining thinking patterns and techniques. The application of tools and scenarios is an urgent problem to be solved in current development (Ahmed and Ali, 2020) and (Aksu, 2019). Fig. 1 shows the statistics of involvement in employing various tools (Liang and Zhikui, 2020). The survey was conducted on 1429 respondents during the pandemic.

Fig. 2 shows the share of digital tools used for various activities (Zhao et al., 2021).

Data mining finds its application in many fields like healthcare,

environment analytics et c. This focus of on developing, applying and researching the automatic methods to extract patterns form large repositories of educational data, which is very hard to tap. DM had a very great lineage and its role in the educational settings which has heterogeneous type of data should not be undermined. AS DM is capable of analyzing any type of data generated from the information system. The data that could be handled by the DM based systems.

The first type of open-source data mining tools represented by WEKA, Rapid Miner, etc., are written in JAVA language, support a variety of standard data mining tasks, and provide the implementation of some scalable data analysis and mining algorithms (Garg, 2020). The second type of enterprise-level big data visualization analysis platform represented by SmartBI and Alibaba Cloud has mature functions such as reports, data visualization, and data mining, helping students experience the commercial application and practice of data mining technology (Rong et al., 2021) and (Lasfeto and Ulfa, 2022). The benefits of digital education are integrated into the virtualized computing system's innovative learning platform. It fixes the issues with dual-mode trials and a lack of resources. Rapid implementation, small expense, and realistic supervision are some of its qualities (Yeye et al., 2022). People use the Internet significantly more often and are well informed. They

E-mail address: [zzb33@zjsru.edu.cn](mailto:zzb33@zjsru.edu.cn).

<https://doi.org/10.1016/j.jksus.2023.102964>

Received 9 February 2023; Received in revised form 9 October 2023; Accepted 20 October 2023

Available online 29 October 2023

1018-3647/© 2023 The Author. Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

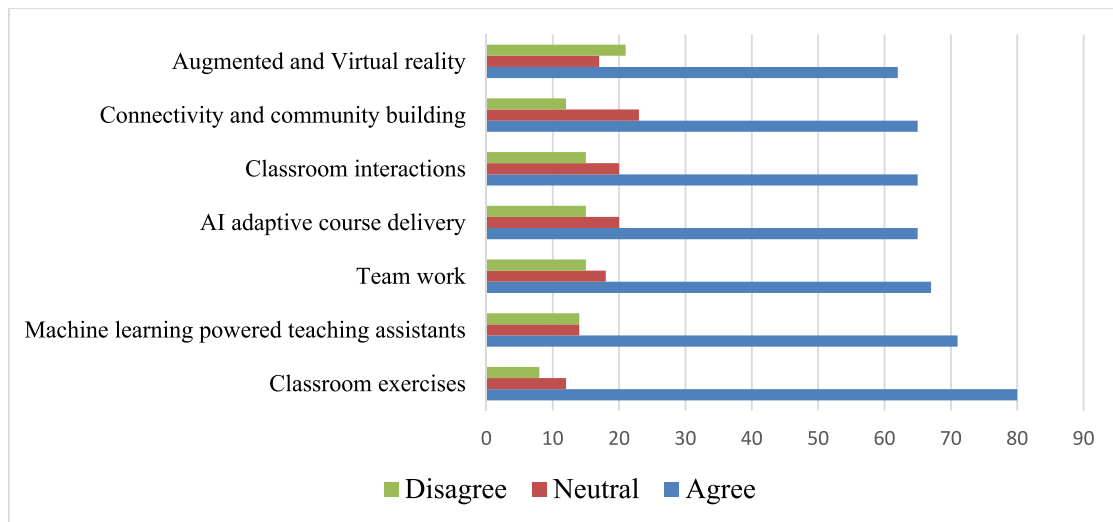


Fig. 1. Statistics in adopting tools.

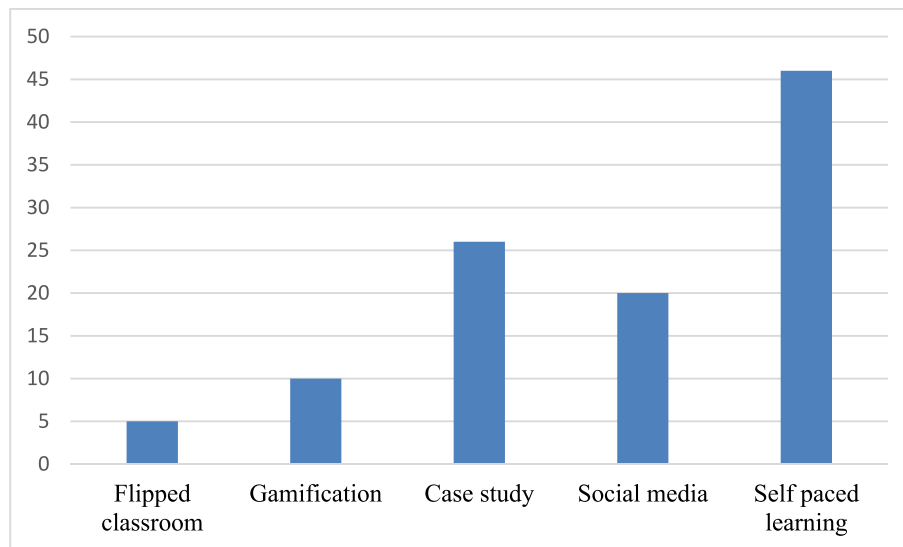


Fig. 2. Usage of digital tools for various activities.

have such a large impact on a variety of organizations as well as a strong effect on global popular opinion (Wu, 2022).

It addresses the system’s programming environment and fundamental technological tools as well as the special characteristics of this component. It displayed the system’s major functional components’ performance mechanism. Using samples and testing tools, it evaluated the technology. The platform’s operational status also evaluates the platform’s energy capability and functionality effectiveness.

**2. Overall Information’s**

Lv Shengping et al. (Mengjuan et al., 2019) build an open and extensible data mining prototype system and developed a teaching demonstration and student practice platform. Through literature review, it is found that most of the current data mining courses rely on third-party software or platforms to carry out experimental, lacking an interactive programming environment, which leads to a disconnect between the introduction of algorithm principles and practice; and the few self-developed experimental platforms do not provide full support for students’ experiments. Information platforms are used by users to view digital movies and carry out group activities. Additionally, IT systems

improve user-to-user interactive communications. Data mining methods are used to evaluate multi-media video patterns to examine customer monitoring behavior at significant moments in time. Users’ habits of viewing videos were examined using data mining techniques in converging IT settings. The process is tracked and recorded, and there is a lack of corresponding teaching feedback and guidance (Liu et al., 2021). Shahabaz and Afzal (Pin et al., 2019) enhanced the effectiveness and adaptability of education-related big data mining technologies utilizing digital twins by resolving the issue of constant contact with distant resources using the features of the database information needed by students. Having simultaneous three-dimensional actual surveillance and wireless remote capabilities depending on a three-dimensional virtual CNC interface is the aim of the virtual interactive environment of the digitally twin-based CNC platforms. Learners can view the training resources before that class and professors discuss them. The instructor may address the issues raised by the learners during class using their observations from evaluating the instructional material. Facebook’s online community gives users access to instructional materials and a variety of activities, including information sharing, evaluating educational resources, and setting shared goals. Zhao (Mihaescu and Popescu, 2021) analyzed the influencing aspects via the decision

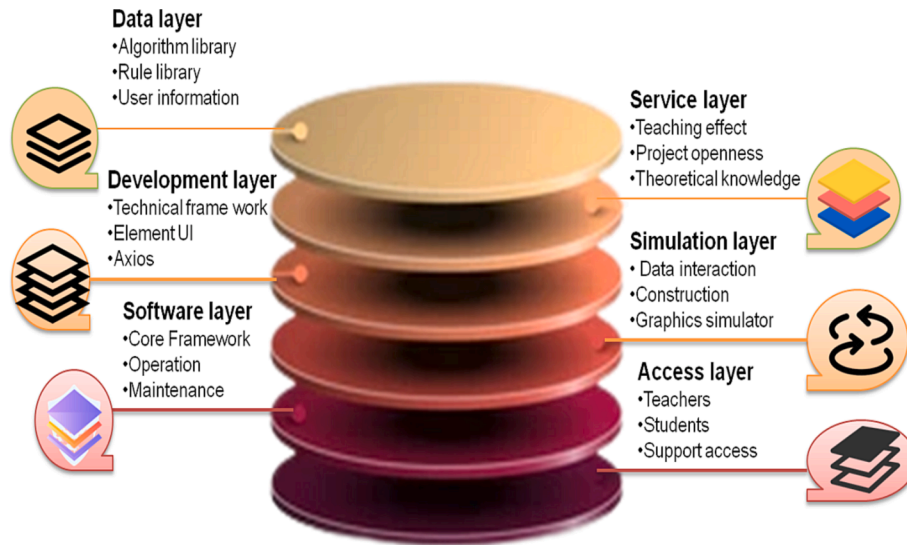


Fig. 3. Data mining flowchart.

tree depending on the mining findings; teachers were able to effectively understand the students' learning circumstances. The aforesaid combination of instructional information and the development of an education model has the potential to greatly enhance digital resources. Finding themes that are similar to and different from each other in the two disciplines as they develop is one way to create clarification, regularity, and integrity around the two sectors. Li et al. (Mody and Bhoosreddy, 1995) employed robots to collect and store data about public perception while introducing the fuzzy sets theory to the system. Data pretreatment is then completed using dispersed textual quantization, selection of features computational resources, and textual representation edge systems. The cross-dimension mining system of popular perception in virtual learning data is built by the findings of the dimensional correlation matrix and preprocessing, using the cloud computing platform as the core. Mihaescu and Popescu (Salihu and Zayyanulyya, 2022) intended to provide a brief and useful summary of the largest popular access to the public data sets, including the EDM goals, employed methods, empirical results, and key conclusions that accompany it. There seem to be three different categories of data resources, according to their research: well-known sources of data, datasets utilized in EDM contests, and independent EDM databases. The visualization of data mining is a process for processing and displaying enormous amounts of data. The impact of techniques and computational paradigms on specific data storage and processing transfer needs. To support big data mining in cloud systems, there are big data mining methodologies that deal with cloud-compatible issues and computational methods (Roger et al., 2020).

Based on the above analysis, this research analyzes the key nodes and steps of data mining, divides the system function modules, optimizes the system performance scheme, builds a system development framework, and designs and implements a Python-based data mining virtual simulation experiment in the teaching system. Based on the system, a series of comprehensive experimental cases are designed in simple language, combining practical application problems with data processing and analysis. Through personalized interactive learning and real-scene experimental training, students gain the whole-process ideas and methods of data mining to solve practical problems.

### 3. System requirements analysis

#### 3.1. Design goals

The overall goal of the system is to provide a virtual simulation

experiment platform for the experiment of data mining courses. The design goals are as follows:

- (1) Help to deeply understand the whole process of data mining, be proficient in using Python-based big data processing and analysis tools, master various functional modules such as data collection, preprocessing, analysis and visualization, and conduct demonstrations driven by the actual solution of application problems analyze.
- (2) By restoring real application scenarios, designing challenging comprehensive experimental projects.
- (3) Virtualization, software packaging, and virtualized storage capacity are the components offered by virtual simulation experiment technologies. These can fulfill the requirements of instructional delivery in computer specialties. Every practice must rationally layout laboratory modeling material for the simulated space by its particular curriculum.
- (4) A movable representation of apps and requirements may be created using the free template Virtual machine process occurring

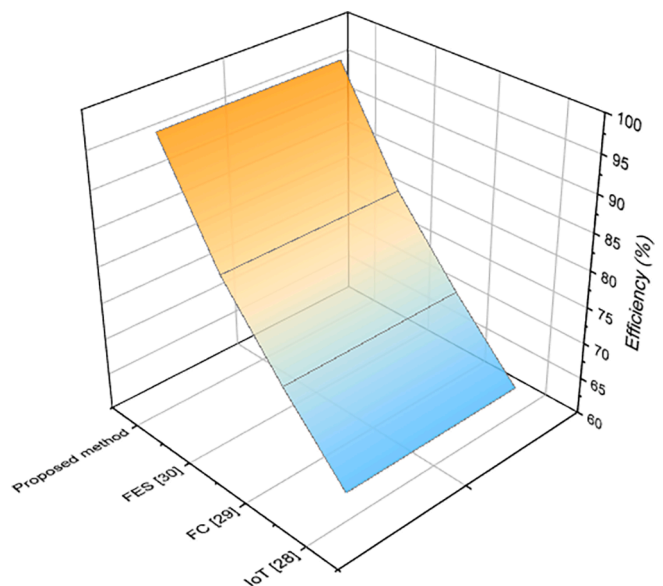


Fig. 4. Overall system architecture.

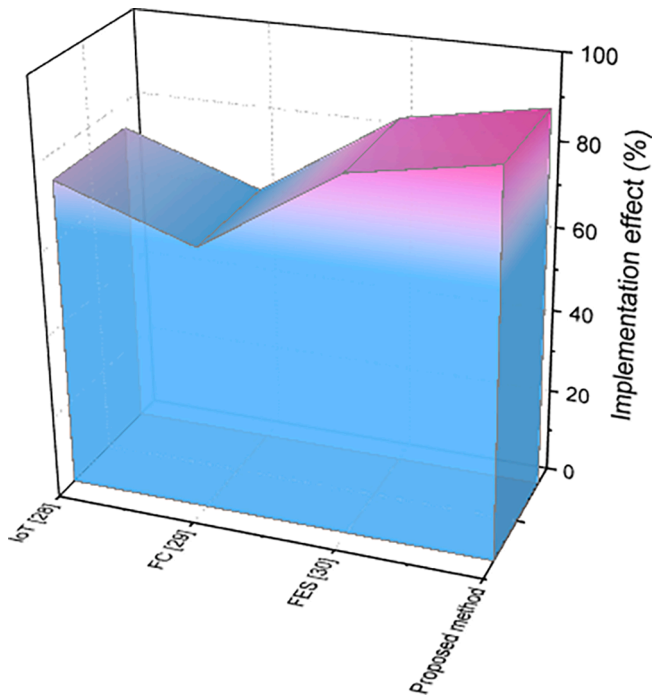


Fig. 5. Main page implementation effect.

technology and distributed on any widely used Linux or Windows computer. Docker, an online platform for creating, distributing, and executing programs, drastically lowers the time among programming languages and has it executed in a manufacturing process by separating implementations from resources.

### 3.2. Process requirements

According to the specific design goals, combined with the data

mining process and the main tasks of each link, the business process is further refined. As shown in Fig. 3, it includes data acquisition, data preprocessing, data analysis, and data visualization.

The main task of data collection is to obtain data sources through local reading or code capture, and are required to have a certain degree of mastery of relevant toolkits for data collection, such as Pandas, Selenium, and Beautiful Soup. The main task of data preprocessing is to process raw data to improve data quality, thereby improving the execution efficiency of data mining modeling and the accuracy of analysis results (Shahabaz and Afzal, 2021). Students are required to master data preprocessing methods based on Pandas and Numpy, and be able to flexibly select data preprocessing techniques based on business requirements and data types. Common techniques include data cleaning (filling, replacement, deduplication, discarding), data transformation (classification, sorting, dimensionality reduction), data reduction (standardization, discretization), feature extraction (TF-IDF, chi-square test, recursion), etc Mishra et al. (Hong et al., 2021).

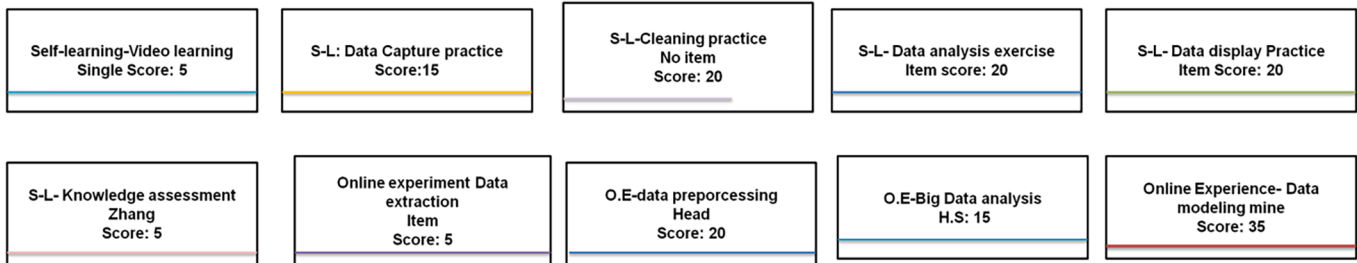
Data mining analysis can be divided into descriptive analysis and predictive analysis according to the task (Zhenwu, 2018). The main task of descriptive analysis is to extract and summarize the patterns of potential associations (correlations, trends, clusters, trajectories) in the data, to quickly grasp the main characteristics and information of the data (Wt et al., 2021). Commonly used descriptive analysis methods include association analysis, cluster analysis, principal component analysis, keyword extraction, etc. The main task of predictive analytics is to discover valid and valuable information by predicting the variable value of a specific attribute based on the variable value of other attributes through statistical and machine learning algorithms. Commonly used predictive analysis methods are classification models (nearest neighbor algorithm, decision tree, vector machine, Bayesian network, etc.) and predictive models (regression, neural network, random forest, etc.) (Torniainen, 2020) Students are required to master common machine learning algorithms based on sklearn (Verboven et al., 2012).

The main task of data visualization is to map the processed data information into visual elements, to achieve the purpose of displaying the results efficiently, concisely, and powerfully so that the details in the

Experimental score Statistics:

Experimenter: Liu pei Class: E-Commerce class 201 Student number: 202002014618

Self learning Score: 85 Online Experiment Score: 80 Total Score: 82



Experimental evaluation:

Please enter the experiment Statistics

0/200

Submit

Fig. 6. The effect of knowledge assessment.

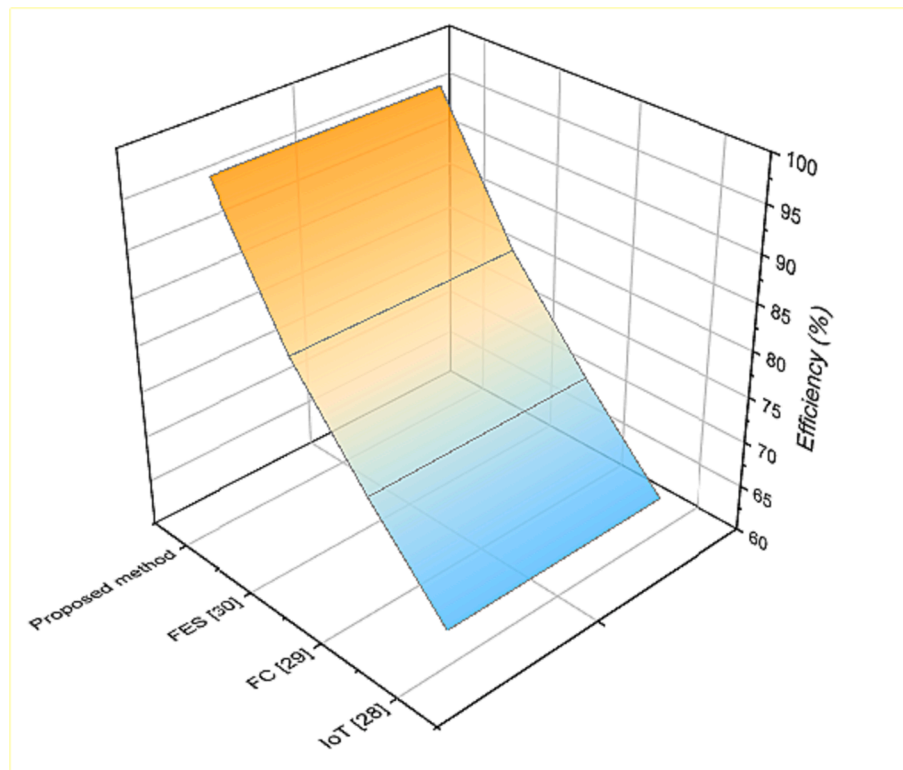


Fig. 7. The effect of the interactive programming practice environment.

data can be more easily mined. Students are required to have a certain degree of mastery of Python-based graphics drawing toolkits, such as Matplotlib, Seaborn, etc. Zhao et al. (Gerretzen, 2015). Common visualization types include text data visualization (word frequency, semantics, geometry), multidimensional data visualization (icons, coordinates), time series data visualization (trend, interval), geospatial data visualization (point, area), etc. Li (Saeys, 2019).

### 3.3. Functional requirements

Combined with the design goals and business requirements, the system function modules are further refined to ensure the reliability and rationality of the system design. Fig. 2 is a functional structure diagram, including three modules: online experiment, and experiment report. The autonomous learning module builds a personalized interactive learning environment, which is divided into two parts: theoretical learning and programming exercises. It adopts the experimental teaching method of autonomous and inquiry-based learning to transform relevant theoretical principles into practical content that can be dynamically interacted with in real-time space (Engel, 2013). The programming practice adopts an online interactive programming environment, covering the practical operation types of the whole process of data capture, preprocessing, analysis and visualization, so that students can write code by themselves, observe the running status and results of the program through the WYSIWYG method, and can help students comprehensively exercise their programming skills and lay a solid foundation for completing complex and challenging comprehensive experimental projects.

The online experiment module highly restores the application scenarios of multiple industries and integrates data analysis and processing technology into practical application problems. The experimental projects include user portrait description and value analysis, user repeat purchase prediction, credit card score modeling analysis and operator user churns prediction. An experimental introduction, a completion part, and a challenging part are designed for each project (Mishra et al., 2020). The introduction to the experiment includes the experimental

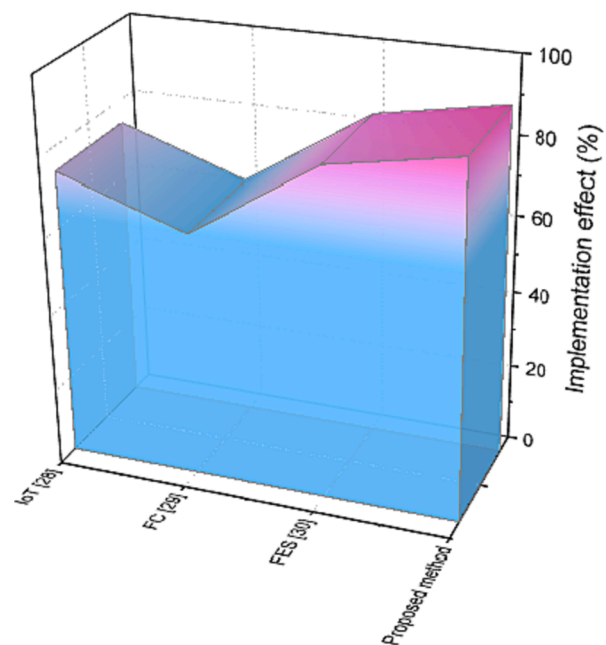


Fig. 8. Efficiency.

background, purpose, methods, procedures, and assistance. The experimental content of the complete part can be completed by the basic model that students have learned in the classroom. Each step of the experiment has targeted result feedback, and the program can be debugged and output interactively in real-time, and the knowledge points can be completed in the process of trial and error and exploration, and consolidation process.

The experimental evaluation module uses an interactive dynamic

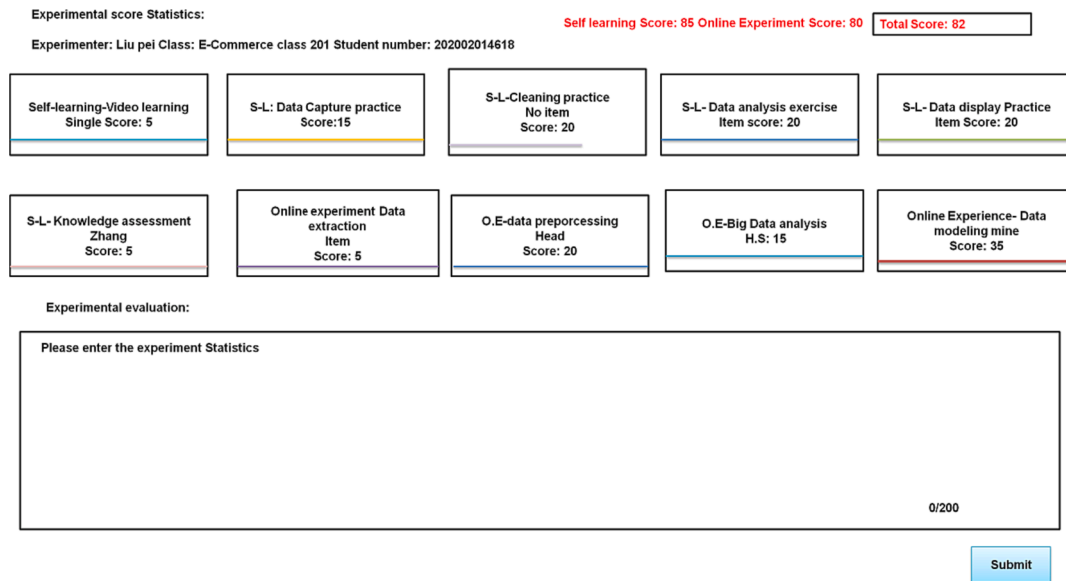


Fig. 9. Data analysis implementation effect.

evaluation method, which is divided into self-learning evaluation and online experimental evaluation. Self-directed learning evaluation records the learning process of students in self-directed learning modules, such as video learning time, knowledge assessment results, and the amount of code written in programming exercises. The completion part of the online experiment evaluation sets the weight and score for each experimental step, the system can automatically judge the score, and the challenging part objectively and comprehensively evaluates the completion of the students' experiments with scientific evaluation indicators according to the results submitted by the students and assesses the student's performance.

### 3.4. Security requirements

The system provides a corresponding solution to the security of data transmission. Generate a token based on the user name or user id, combined with the user's IP or device number. When requesting the backend, the backend obtains the token in the http head, checks whether it is legal (consistent with the records in the database or Redis) and stores it in the database/redis (Mody and Bhoosreddy, 1995) when logging in or initializing. Both the client and the server store a secret key, each transmission is encrypted, and the server decrypts according to the secret key.

### 3.5. Performance requirements

In this platform, the java background runs the python script. To achieve a quick preview of the results, the client directly interacts with the server, and multithreading is used in the interaction with python to ensure execution efficiency and execution stability. Use synchronized code blocks to eliminate redundant code and Enable server caching. Directly access the static resource result through the domain name IP. Minimize disk access when doing I/O operations. Use Buffered Reader to create buffers to release large object resources as soon as possible. Use Redis for system data caching to improve the concurrent performance of the entire platform.

## 4. Overall system architecture

The system is developed using service-oriented software architecture, as shown in Fig. 4. It is divided into the data layer, development layer, software layer, service layer, simulation layer, and access layer.

Each layer provides services for its upper layer until the completion of the specific virtual experiment teaching Construction of the environment.

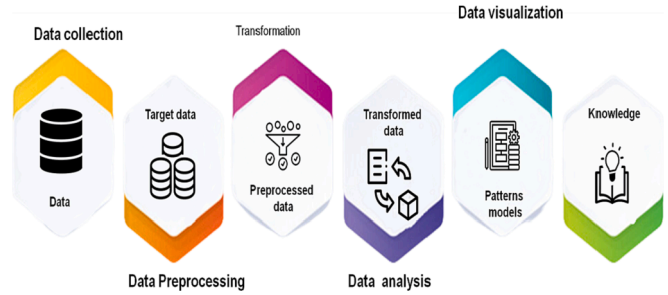
- (1) Data layer. The system involves various types of industry data and mining algorithms and sets up an experimental course library, a typical experiment library, a standard answer library, an algorithm library, a rule library, and a user information library to store and manage the corresponding data.
- (2) Development layer. The development layer is the technical framework of the system implementation, providing a light-weight and easy-to-use interactive interface, a virtual scene of multi-industry data mining applications so that students can learn and experiment anytime, anywhere. The development layer includes: web front-end architecture, covering vue.js, ElementUI, Axios, Vue-router, and other technologies; the back-end adopts Flask architecture, covering programming languages such as Ipython, Numpy, Scipy, Matplotlib, Pandas, etc. The development tools implemented by the system adopt Microsoft Visual Studio, Sourcetree, GitHub, Visual 3D, Sketch, and so on.
- (3) Service layer. The service layer provides general support components for the virtual experimental teaching environment, including theoretical knowledge learning, experimental resource management, experimental report management, teaching effect evaluation, intelligent guidance, automatic evaluation of experimental results, automatic upload of experimental data, project openness, etc. Corresponding integration interface tools are provided so that the system can easily integrate third-party virtual experiment software into unified management.
- (4) Simulation layer. The simulation layer mainly provides scene interaction and data interaction, including scene construction, character construction, editor construction, and graphics rendering, provides a general simulator, and finally provides a formatted output of experimental result data for the upper layer.
- (5) Access layer. The access layer can be assigned different permissions according to different user levels, including administrators, teachers, and students, and supports access to mainstream browsers such as Chrome, Firefox, and Edge.

## 5. System implementation

The main interface of the system is shown in Fig. 5, which mainly

**Table 1**  
Completion of experimental design.

Module	Target	Content	Method	Score
Data extraction	Read data information	After completing the reading of industry data, the read content is classified according to different content fields of the industry. After completion, the read data will generate a new data table.	Call info(), describe(), value_counts() and other functions to view data information.	5
Data preprocessing	Clean and combine sales flow meter and member information table	Handle incomplete and logically wrong problems in data, such as data format conversion, missing value, duplicate value processing, outlier removal, etc.	Call drop(), merge(), drop_duplicates(), etc. functions to process the data.	20
Data analysis	Statistical analysis and visualization, combined with data exploration to building features	The age distribution, the consumption power of different age groups, consumption situation of different genders, number of orders and consumption ratio of members/non-members, sales trends of shopping malls, trends of members' annual and monthly consumption, and consumption conditions at different times.	Perform exploratory analysis on existing data tables through histograms, scatter bubble charts, line charts, pie charts, and more.	20
Data modeling	Create member user portraits	According to the label rule generation method, the constructed members' age, gender, point level, consumption frequency, consumption level, average value per order, shopping product preference, and other labels are divided into user basic feature labels, business feature labels, and interest feature labels. three parts.	Extract relevant data, count the number of member user tags and save them in the dictionary, and use the word cloud library to draw a word cloud of user portraits.	25
	Use purchase data to segment users and build clustering models	Through K value optimization, the K-means algorithm is used for cluster analysis. Combined with the business, evaluate and analyze the characteristics of a certain group by comparing the size of each characteristic among the groups, and summarizing the superior and weak characteristics of each group.	Through the RFM model introduced in the theoretical class, the key features of identifying user value are extracted and a new model is constructed.	30



**Fig. 10.** Experimental score statistics implementation effect.

**Table 2**  
Comparison of existing and proposed method for efficiency.

Sl.No.	Methods	Efficiency (%)
1	IOT (Yan and Jia, 2022)	64
2	FC (Wenmei et al., 2021)	73
3	FES (Zhang, 2021)	83
4	Proposed method	97

includes independent learning, online experiments, and experimental evaluation modules. To improve the speed of concurrent page access and display of self-learning information (video viewing time, coding practice completion code, knowledge assessment average score), online experiment completion information (experiment completion progress, experiment completion duration), the main page content is added to the cache. For requests, the data is called from the Redis cache first, and the third-party interface is not called in the cache.

**5.1. Self-directed module**

This module mainly implements knowledge assessment and interactive programming practice environment, and provides a series of learning resources, videos, classic algorithms, scientific literature, etc., to help students understand the basic knowledge of data mining, data mining algorithms, and be familiar with the experimental stage. Basic theory. Fig. 6 shows the realization interface of knowledge assessment. The system has a built-in question bank for judgment and multiple-choice questions. Before the assessment, some judgment questions and multiple-choice questions are extracted and combined. Through the combination and sorting of the two system question banks, the knowledge assessment module becomes more diverse.

Fig. 7 shows the implementation interface of the programming exercise, where students can write code, run, debug, and view the results online. At the same time, the system has built-in a set of standard data capture, data preprocessing, and data analysis and data visualization codes. Students can practice multiple times in modules to become familiar with the realization method of the whole process of data mining. Taking the data capture practice module as an example, students with weak foundations can obtain data content of different periods, different types, and different dimensions according to their ideas by modifying the relevant variable content in the standard code and the filtering conditions of data capture, and then The relevant data content is aggregated into a new data table, and subsequent data operations are performed based on the currently generated data table. Students

**Table 3**  
Comparison of implementation effect with existing and proposed method.

Sl.No.	Methods	Implementation effect (%)
1	IOT (Yan and Jia, 2022)	74
2	FC (Wenmei et al., 2021)	63
3	FES (Zhang, 2021)	84
4	Proposed method	90

practice many times until they can write the code themselves.

### 5.2. Experiment module

A total of 5 comprehensive experiments are designed in this module, and each comprehensive experiment includes three parts: experiment introduction, completion experiment, and challenging experiment. The following takes the experiment of "Mall Member User Portrait Depiction and Value Analysis" as an example to introduce the implementation method of the online experiment module.

The introduction to the experiment includes the experimental background, purpose, methods, procedures, and assistance. The application scenario is to draw user portraits and finely divide members through the processing of real shopping mall operation data and member information data, to formulate corresponding marketing strategies for different groups, thereby increasing the sales profits of shopping malls. The capacity to do a task with little to no waste, effort, or energy is referred to as efficiency. Being effective is using the resources you have as effectively as possible to attain your goals. Simply put, if there is no wastage and all procedures are streamlined, something is efficient. Fig. 8 shows the comparison of efficiency for existing and proposed methods (See Fig. 9.).

The completion experiments of this comprehensive experiment include data extraction, data preprocessing, data analysis, and data modeling. The objectives, contents, methods, and assignments of each module are shown in Table 1. The system evaluates the completion of students' experiments by setting a scoring model for the achievement of goals for each module. For example, the data preprocessing module requires students to clean all dirty data (5 points), deal with garbled data (5 points), duplicate data (5 points), and Null data (5 points), 10 min time limit, 20 points for reaching the goal. Completion experiments are based on an interactive online programming environment.

The challenging experiment requires to expand the RFM model, submit the experimental report (including improvement plan, difficulty analysis, and performance index comparison analysis), and code the data package online after completion. Teachers can view the submission records, review the experimental report and run the program package.

### 5.3. Experimental evaluation module

Experimental information. In the operation of the experimental process, all the operation steps involved in scoring are counted, and the total steps, correct steps, and other data of each module are displayed to the students for viewing, which is convenient for the statistics of the experimental scores. Fig. 10 shows the implementation effect diagram of the experimental score statistics, showing the sub-item scores of each module. Click to view the detailed step scores.

## 6. Discussion

Depending on the research outcome, data mining may help researchers studying teaching methods quickly identify the users of executive function categories and engage with technological tools to overcome problems. A systematic teaching assessment system must be established to conduct estimates and an examination of the current state of quality process for the management department of colleges and universities, which is responsible for monitoring and counting the teaching quality of teachers. A significant proportion of colleges and universities have amassed a substantial quantity of teaching qualitative data over several decades of quality information response. Hence, we analyzed the virtual simulation experiment teaching platform for the teaching and experiment of data mining courses. The process creates drawbacks connected with access, usage, monitoring, and ownership of data, in addition to standardization that comes under "profiling" rather than personalization, and these disadvantages are compounded by the problem of interoperability on IoT highways (Yan and Jia, 2022).

Because of this issue, we suggest our proposed method to resolve it. Complexity and an inability to recover from database corruption are two disadvantages of fuzzy clustering (FC) (Wenmei et al., 2021). No matter which node in the fuzzy cluster (FC) is responsible for operating the service, the cluster will always utilize the same IP address for the Directory Server and the Directory Proxy Server. Techniques based on fuzzy expert system (FES) (Zhang, 2021) is effective when applied to the resolution of difficult, ill-defined situations that are characterized by the unpredictability of the environment and the fuzziness of the information. The management of uncertain and imprecise knowledge is made possible by fuzzy logic, which also serves as an effective foundation for reasoning. Table 2 shows a performance of existing and proposed method for efficiency. When compared to existing method, our proposed method provides 97 % of efficiency.

Table 3 represents the comparison of implementation effect with existing and proposed method. Our proposed method delivers 90 % of the implementation impact, which is an improvement over the previous approach and compares favorably to the existing way.

## 7. Conclusion

Data mining and visualization is a key teaching project of Artificial Intelligence (AI) and big data technology, which is of great significance to the teaching of data analysis. Given the current data mining algorithms are numerous and complex, technical difficulty is high, experiments are comprehensive, design parameters are highly correlated and coupled, and it is difficult for students to experience the entire business process in combination with reality. This work proposes, designs and implements a data mining virtual simulation experiment teaching system with a personalized Interactive learning environment. This also included AI based real scenic creation and the experimental training based on the multiple industry data using the DM algorithms for gaining deeper understanding and application to systematic comprehensive experimental project. This AI based environment, helps the students to leverage and apply what they have learned and realize the transition from imparting knowledge to train students independently by analyzing the problem-solving skills. The results indicates that the proposed system was able to improve students' participation and autonomous learning ability. The goal of the work is broken down into many predictor levels using modeling and analysis techniques. The system social attitudes score ratio can be obtained based on the structural model until the similarity measure of the quantitative measurements' particular weight values at multiple levels, and the disclaimer level to be stimulated is calculated based on the limit duration equivalent to the impact factor. It is hoped that the experimental teaching system proposed in this paper can provide a useful reference for other colleges and universities in the training of data mining talents.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The article was supported by the 2022 Ministry of Education's Industry-University Cooperation and Collaborative Education Project "Research on the Application of Project-based Learning in Experimental Courses Supported by Smart Technology" (Project No.: 220502054240108) and the 2023 Zhejiang Modern Service Industry Research Center Open Fund Project Grant (Project No.: SXFJY202302).



## References

- Ageed, Z.S., Zeebaree, S.R., Sadeeq, M.M., Kak, S.F., Yahia, H.S., Mahmood, M.R., Ibrahim, I.M., 2021. A comprehensive survey of big data mining approaches in cloud systems. *Qubahan Acad. J.* 1 (2), 29–38.
- Ahmed, B., Ali, A., 2020. Usage of traditional chinese medicine, western medicine and integrated Chinese-western medicine for the treatment of allergic rhinitis. *Official Journal of the Zhende Research Group* 1 (1), 1–9.
- Aksu, D., 2019. An analysis program used in data mining: WEKA. *Journal Measur. Evaluation Educ. Psychol.-EPOD* 10, 80–95.
- Engel, J., et al., 2013. Breaking with trends in pre-processing? *TrAC Trends Anal. Chem.* 50, 96–106.
- Garg, H., 2020. Digital twin technology: revolutionary to improve personalized healthcare. *Sci. Prog. Res. (SPR)* 1.1.
- Gerretzen, J., et al., 2015. Simple and effective way for data preprocessing selection based on design of Experiments. *Anal. Chem.* 87 (24), 12096–12103.
- Hong, S., Hongxia, D., Xiaoye, C., 2021. Visualized teaching and experimental scheme of data mining technology using cloud computing. *Lab. Res. Exploration* 1 (1), 115–119.
- Lasfeto, D.B., Ulfa, S., 2022. Modeling of online learning strategies based on fuzzy expert systems and self-directed learning readiness: the effect on learning outcomes. *J. Educ. Comput. Res.*, 07356331221094249
- Liang, Z., Zhikui, C., 2020. Design and implementation of big data algorithm library teaching experiment platform. *Experimental Technol. Manage.* 37 (6), 197–201.
- Liu, J., Shi, G., Zhou, J., Yao, Q., 2021. Prediction of college students' psychological crisis based on data mining. *Mobile Informat. Syst.*
- Mengjuan, L., Guichuan, Z., Yao, L., et al., 2019. Design and practice of challenging comprehensive experiments for data mining courses. *Experimental Sci. Technol.* 17 (1), 85–88.
- Mihaescu, M.C., Popescu, P.S., 2021. Review publicly available datasets for educational data mining. *Wiley Interdisc. Rev.: Data Mining Knowledge Discov.* 11 (3), e1403.
- Mishra, P., Biancolillo, A., Roger, M.J., et al., 2020. New data preprocessing trends based on an ensemble of multiple preprocessing techniques. *TRAC Trends Analytical Chem.* 132, 116045.
- Mody, R.N., Bhoosreddy, A.R., 1995. Multiple odontogenic keratocyst: a case report. *Ann. Dent.* 54 (1-2), 41–43.
- Pin, L., Siyu, P., Jia, X., et al., 2019. Network traffic classification algorithm based on machine learning. *J. Guangxi Univ. (nat. Sci. Ed.)* 44 (6), 1651–1657.
- Roger, J.-M., Biancolillo, A., Marini, F., 2020. Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy. *Chemometr. Intell. Lab. Syst.* 199, 103975.
- Rong, G.u., Han, Y., Zhaokang, W., et al., 2021. Design of big data online training platform and comprehensive experimental course system. *Experimental Technol. Manage.* 38 (7), 201–207.
- Saeyns, W., et al., 2019. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: a review. *Postharvest Biol. Technol.* 158.
- Salihu, S.O., Zayyanulyya, 2022. Assessment of Physicochemical parameters and Organochlorine pesticide residues in selected vegetable farmlands soil in Zamfara State, Nigeria. *Sci. Prog. Res. (SPR)* 2 (2).
- Shahabaz, A., Afzal, M., 2021. Implementation of high dose rate brachytherapy in cancer treatment. *SPR* 1 (3), 77–106. <https://doi.org/10.52152/spr/2021.121>.
- Torniainen, J., et al., 2020. Open-source python module for automated preprocessing of near infrared spectroscopic data. *Anal. Chim. Acta* 1108, 1–9.
- Verboven, S., Hubert, M., Goos, P., 2012. Robust preprocessing and model selection for spectral data. *J. Chemom.* 26 (6), 282–289.
- Wenmei, Z., Binbin, Q.i., Wenxiang, F., 2021. Data-driven teaching behavior analysis: current situation, logic and development trend. *J. Dist. Educ.* 39 (1), 84–93.
- Wt, W.u., Yj, L.y., Az, F., et al., 2021. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil. Med. Res.* 8 (1), 44.
- Wu, C., 2022. Effect of online and offline blended teaching of college english based on data mining algorithm. *J. Inf. Knowl. Manage.*, 2240023
- Yan, J., Jia, H., 2022. Construction of marine professional virtual simulation experiment platform based on internet-of-things technology. *Secur. Commun. Netw.*
- Yeye, L.i., He, L.i., Wang, S., et al., 2022. Construction of online review knowledge graph based on multi-source heterogeneous data mining. *Informat. Sci.* 40 (2), 65–73.
- Zhang, Z.B., 2021. Research on learning evaluation of online general education course based on BP neural network. *Comput. Intell. Neurosci.* 3570273.
- Zhao, J., Guan, H., Lu, C., Zheng, Y., 2021. Evaluation of teachers' educational technology ability based on fuzzy clustering generalized regression neural network. *Computational Intell. Neurosci.*
- Zhenwu, W., 2018. Discussion on the experimental teaching method of "data mining" course based on Alibaba Cloud big data platform. *Lab. Res. Exploration* 6 (6), 192–196.