Journal of Intelligent Computing

DLINE JOURNALS

Print ISSN: 0976-9005
Online ISSN: 0976-9013

JIC 2025: 16 (1)
https://doi.org/10.6025/jic/2025/16/1/10-18

Building an Intelligent Education Model for Student Profiling Based on Big Data Algorithms

Yubao Shen
Department of Intelligent Science and Technology
Hefei Preschool Education College
Hefei, Anhui. China
ybshen@hfpec.edu.cn

ABSTRACT

The development of big data technology has driven the pace of teaching innovation and reform. In the information age, education emphasizes personalized and comprehensive development of students more than ever before. This paper combines big data algorithms to construct an intelligent education model for classroom student profiling. The model leverages big data mining algorithms to discover the correlations in student behavior data. Using classification algorithms based on multi-frequency patterns, the model classifies student behavior data and constructs multi-frequency pattern trees for students with different academic performance, reflecting differences in their learning behavior characteristics. Experimental results demonstrate that applying the intelligent education model based on big data algorithms can effectively provide teachers with comprehensive and accurate feedback on student behavior characteristics, helping students understand their learning situations and enabling targeted personalized teaching, significantly improving students' learning quality and efficiency.

Keywords: Big Data Algorithms, Classroom Education, Student Profiling, Intelligent Education Model

Received: 18 October 2024, Revised 28 December 2024, Accepted 8 January 2025

1. Introduction

The concept of "Internet+ education" has promoted the pace of educational reform. The rapid development of next-generation information technology has provided more development directions for educational reform and changed how teachers teach and students learn. Constructing an information-based and intelligent education system has become a trend in educational innovation. In educational innovation and reform, the emphasis is on smart and personalized education management, diversified and shared educational resources, precise student management, and the cultivation of personalized and diversified talents that meet the needs of the times[1]. Different students have different learning behavior characteristics and learning tendencies. Traditional educational models focus more on the overall learning situation of students, thereby neglecting the issue of individualization for each student. Regarding student management, it isn't easy to provide targeted help to students to improve their academic performance. In addition, students' learning behaviors, life status, and behavior habits, among other personal factors, can affect their learning outcomes and learning quality. With limited teacher resources, teachers cannot individually analyze the actual learning situations of each student or fully grasp their learning statuses to recommend appropriate learning resources and methods for students[2]. The development of big data technology has changed the traditional teaching model and broken the communication barriers between teachers and students. Big data algorithms can collect and analyze a series of data information from students' online learning, extract valuable data, and gradually construct a learning behavior profile for each student, providing precise information and patterns of student learning behavior for education managers. Based on this information, education managers can effectively intervene in and guide corresponding student learning behaviors[3].

Education data mainly includes structured data and semi-structured data. Significant differences exist in the storage, content, format, and integrity constraints between different data types, resulting in scattered, lack of uniformity, and inconsistent educational data[4]. Therefore, this paper is based on structured student data extracted through data warehousing technology to mine and analyze student learning behavior data and adopt the classification algorithm of a multi-frequency pattern tree to construct a student profiling fusion model, improving the accuracy and effectiveness of student profiling and providing strong data support for teachers to adjust teaching plans and implement personalized teaching[5].

2. Current Research Status of Education Data Mining

Education data is an essential aspect of educational reform and development. However, due to its diversity, complexity in structure, and vast content, comprehensive regular analysis was challenging using traditional data analysis methods in the early stages. The development of big data computing has changed the application scope of education data and achieved the goal of in-depth data mining in education. In the 21st Century Universal and Emerging Education Model Program by the U.S. Department of Education, they proposed the strategy of promoting national education plans driven by science and technology. Therefore, scholars have transformed traditional educational management models through the collection and processing of massive data to achieve a reinforced learning model[6]. Some scholars have proposed the application of big data mining and learning analytics technology in education to address the problems faced in the implementation of teaching management. Based on this, the U.S. National Science Foundation believes that innovative educational concepts will be widely applied and deepened with the development of information technology. The development of education data mining will gradually be promoted in education, helping educators analyze, guide, and research students' learning and providing scientific decision-making information. The development of education data mining has promoted the development of personalized and accurate education concepts. Efficient educational institutions have achieved educational improvements through data analysis technology, effectively helping students improve their grades and enhancing the effectiveness of student management[7]. It has also deepened the analysis and understanding of students' learning by education managers. Regarding the analysis methods of education data, some scholars have proposed the concept of data clustering to address the low level of intelligentization in education data. This concept classifies typical data to achieve efficiency and scope in applying education data in different fields. Other scholars have suggested combining machine learning algorithms for data analysis, researching the problems in teaching based on the correlations among student learning content, methods, behaviors, resources, outcomes, and teaching models[8]. Some scholars have analyzed the relationship between students' daily behavior data and learning performance data using association rules and sequence patterns, presenting the final analysis results through visualization. Currently, many universities have begun to establish student behavior early warning management systems, which select students with financial difficulties based on daily behavior and provide assistance, or analyze students' learning statuses and issue corresponding warnings, thereby promoting student learning [9].

3. Intelligent Education Model for Classroom Student Profiling Based on Big Data Algorithms

3.1 Data Preprocessing and Data Mining Process for Student Profiling

Student profiling, also known as student learning behaviour profiling, involves the virtual operation of behaviour data. Its essence lies in visualizing data models obtained through data association mining. In practical applications, some raw data may lack completeness, accuracy, and consistency or contain outliers, which can affect the results of big data algorithms in data mining. Therefore, before applying big data algorithms to mine and analyze large volumes of data, the raw data must be cleaned to improve data quality and provide a solid data foundation for subsequent processing. The sources of student behavior data mainly include student card information, consumption information, student academic management system information, and library borrowing information, among others [10]. These data can present and describe students' daily and learning behaviors from multiple perspectives. This paper chooses to use Lagrange interpolation to deal with missing data in student information. For example, considering student course grades, if there are N data points, a polynomial of degree N-1 can be obtained that passes through all N grade data points, as shown in formula (1):

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1} \quad (1)$$

Bring the existing student course grade coordinates into the above formula and release the results, as shown in formula (2):

$$\begin{aligned} L(x) &= y_1 \frac{(x-x_2)(x-x_3)\ldots(x-x_n)}{(x_1-x_2)(x_1-x_3)\ldots(x_1-x_n)} + y_2 \frac{(x-x_1)(x-x_3)\ldots(x-x_n)}{(x_2-x_1)(x_2-x_3)\ldots(x_2-x_n)} \\ &+ y_n \frac{(x-x_2)(x-x_3)\ldots(x-x_{n-1})}{(x_n-x_1)(x_1-x_2)\ldots(x_n-x_{n-1})} = \sum_{i=0}^{n} y_i \prod_{j=0, j\neq i}^{n} \frac{x-x_j}{x_i-x_j} \end{aligned} \quad (2)$$

The independent variable is, the dependent variable is, and the approximate value of missing grades is described as.

The validity of outliers in the original data should be judged first. If there is validity, the data should be analyzed directly. Otherwise, outliers should be eliminated. After processing missing values and outliers, consistency processing should be carried out for the complexity and contradiction of data. Not all of the processed data is related to students' learning behaviour. Therefore, it is necessary to reselect the data based on the analysis purpose target related data features and form a new recognition feature dataset. In the process of selecting feature data, the classification information of the data object's attributes should be largely preserved to avoid increasing the dimensionality of features.

In addition, there are significant differences in the magnitude and numerical values of the corresponding data generated by different behaviors of students. This article uses the minimum, maximum normalization method to distribute the behavior data of students with different magnitudes within the same range, as shown in formula (3):

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (3)$$

In the equation, the initial data of student behavior data is represented as, where the maximum value is, the minimum value is, and the normalized behavior data value is described as.

Student behavior data mining is based on knowledge discovery and data mining algorithms to obtain knowledge patterns of student behavior preferences. This process requires analysis of the correlation relationships between random data objects. This article uses Pearson correlation coefficients, as shown in formula (4):

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (4)$$

The correlation coefficient is expressed as in the equation, and its value range is. Different values represent different correlations between students and behavioral data.

The Model selection of the data mining process in this paper is CRISP-DM, which focuses on data application analysis to solve the practical business problems of mining technology.

3.2 Student Behavior Portrait Process and Model Fusion

Student behavior profiling is a data set that layers and classifies the behavior data generated by students according to the corresponding label characteristics through big data algorithm. This set describes the data and semantics of behavior, and can describe the characteristics of behavior based on global significant probability. Continuous iteration based on this feature can enhance the accuracy of the model's judgment on the likelihood of student behavior occurring in scenarios. Therefore, student behavior profiling is not about constructing the average characteristics of all students or individual characteristics that students possess, but rather the group characteristics formed by specific students based on needs. The obtained student behavior portraits can reflect the high probability of students' situational behavioral characteristics, predict their learning behavior, and provide targeted guidance and intervention. As shown in Figure 1, the student behavior portrait Process flow.

In constructing student behaviour portraits, clustering analysis of student behaviour data is required based on features, and frequent pattern-based classification algorithms can obtain frequent patterns between the itemsets contained in the database based on the correlation between objects and data. Based on this article's requirements and actual situation, the frequent pattern algorithm used is the FP growth algorithm, a divide and conquer algorithm based on horizontal data format to obtain frequent item sets. Assuming the transaction set contains things, the two metrics include support, confidence, and improvement. If there is a correlation between the metrics, it can be described. As shown in formulas (5) - (7), the calculation formulas for the three metrics are:

$$S(A \Rightarrow B) = P(A \cup B) \quad (5)$$

$$C(A \Rightarrow B) = P(B|A) = \frac{P(A \cup B)}{P(A)P(B)} \quad (6)$$

$$l(A \Rightarrow B) = \frac{C(A \Rightarrow B)}{S(B)} = \frac{P(A \cup B)}{P(A)P(B)} \quad (7)$$

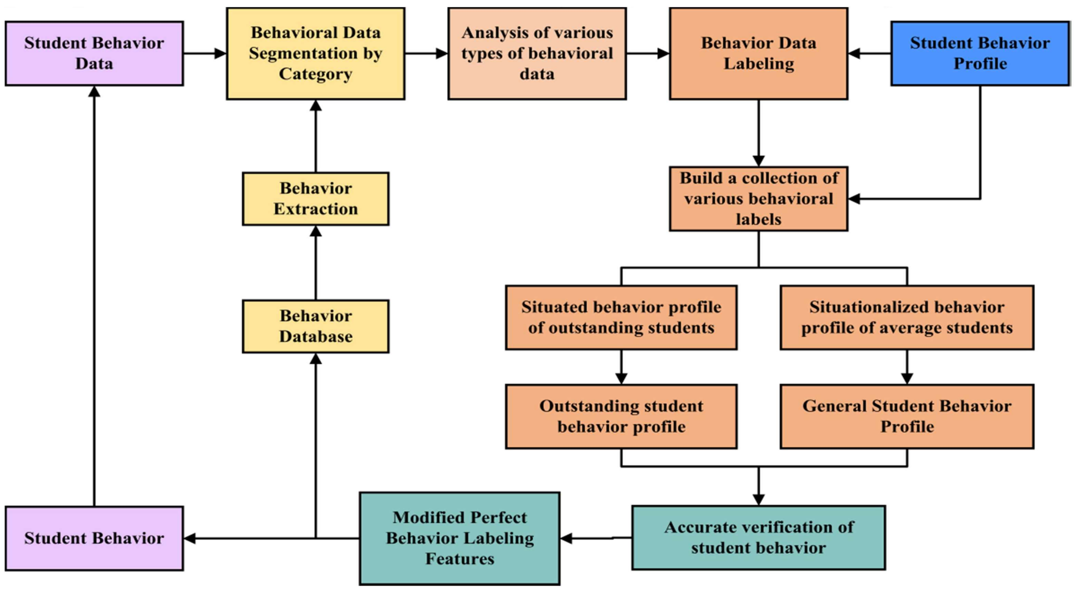

Figure 1. Student Behavior Portrait Process Flow Diagram

In the formula, the support level, the confidence level, and the improvement level are denoted. When the formula indicates that transactions are independent of each other, it indicates a positive correlation between transactions, and vice versa.

Multiple frequent patterns are strongly correlated material rules that meet the minimum support and minimum confidence threshold conditions at the same time. To improve the overall quality of student portraits, this paper will classify the frequent patterns of multiple frequency patterns and form a new fusion model with machine learning integration algorithms from various countries in childhood to identify the behavior characteristics of excellent and ordinary students. The fusion model is mainly divided into three parts. First, the student data set will be mined with Discretization tags, and frequent item sets will be obtained and multiple frequent patterns in the data will be found according to the preset support threshold. Secondly, the numerous frequent patterns are analyzed and strongly correlated material rules are obtained to exclude the multiple frequent patterns that are weak in association and redundant multiple frequent patterns. Finally, a frequent pattern tree of student data is constructed based on multiple frequent patterns, and practical storage and extraction of frequent patterns are achieved through the data tree structure. Its pruning strategy is to record the distribution of Discretization data category symbols of each multi-frequent pattern in the set of frequent patterns that meet the conditions, which is mainly obtained through the deformation structure of the FP growth algorithm. As shown in Figure 2, there is a tree of frequent patterns among students with different grades.

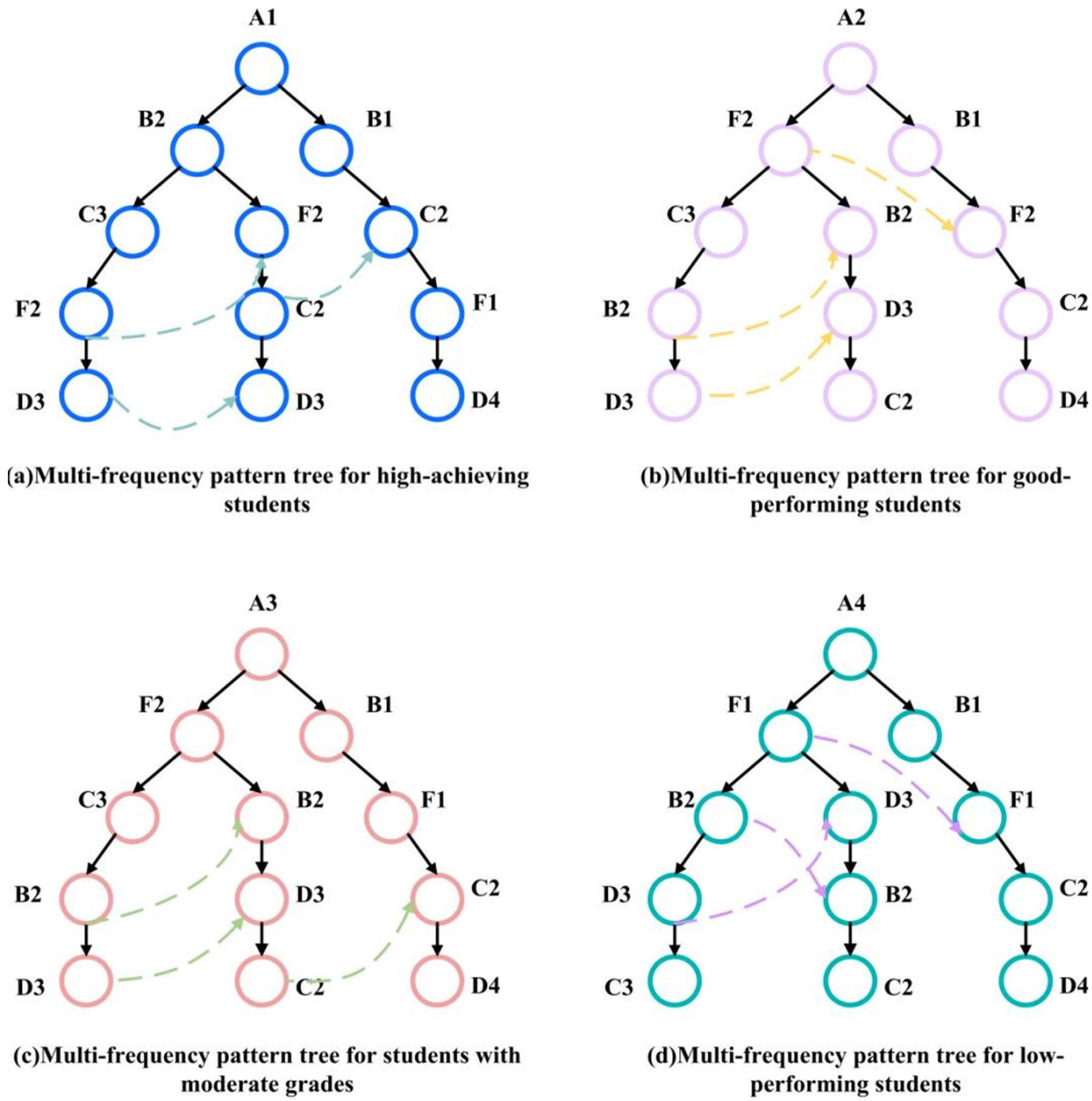

Figure 2. Student Multi-Frequency Pattern Trees for Different Grades

4. Results of the Student Profiling Intelligent Education Model Based on Big Data Algorithms

A comparative experiment was conducted in a randomly selected first-grade class in a secondary school to verify the applicability of the classroom intelligent education model based on big data algorithms. During the experiment, the class was divided into two learning stages: the first half of the semester followed the traditional teaching mode. In contrast, the second half applied the classroom student intelligent education model for personalized teaching. According to the pre-experiment survey results, the overall learning behavior engagement of the students in the class was not consistent, and there were significant differences in behavior. Figure 3 shows the comparison results of students' grades and correct rates for important test points before and after the experiment. The data in the figure shows that after applying the classroom student intelligent education model for personalized teaching, the overall grades of the class students improved significantly. The most tremendous improvement was observed in declarative knowledge, with an increase of over 40%. The most negligible improvement was in procedural knowledge, which remained relatively similar. However, from the standard deviation, it can be seen that the distribution of procedural knowledge scores in the first half of the semester was relatively scattered. In contrast, in the second half of the semester, it became more concentrated. This indicates that the student's overall procedural knowledge was optimised, reducing the number of students with low scores, and the distribution of scores became more concentrated. Figure 3 shows that in the first half of the semester, the students did not perform well in understanding the five important test points, with most of them only mastering two. In the second half of the semester, the students in the class were able to grasp the knowledge of all five test points, and the correct rates improved significantly. Therefore, it can be concluded that the application of the classroom student intelligent education model based on big data algorithms can effectively and accurately construct student profiles and provide teachers with highly accurate data information, helping them improve the quality and effectiveness of personalized teaching and providing targeted assistance to different students to enhance their learning outcomes.

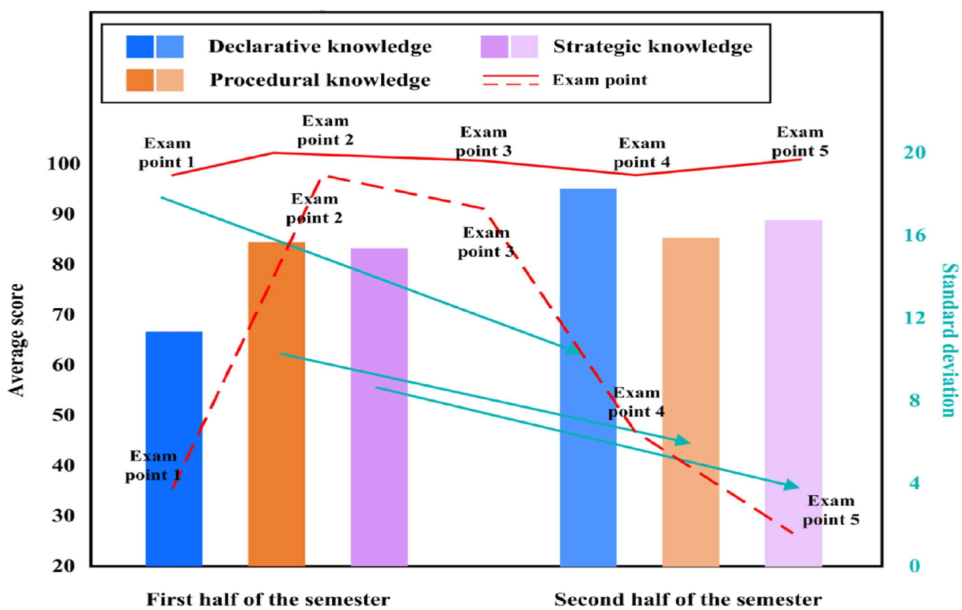

Figure 3. Comparison of Students' Grades and Correct Rates for Important Test Points Before and After the Semester

Figure 4 shows the evaluation results of the class students' satisfaction and applicability with the personalized teaching model based on student profiling. The data in the figure shows that over 90% of the students believe that the intelligent classroom education model based on big data algorithms can accurately construct student profiles and help students and teachers comprehensively understand students' learning behavior characteristics and situations. Furthermore, the personalized teaching approach based on student profiling is better suited to the individual situations of different students, providing targeted assistance to address their learning issues regarding knowledge points, skill operations, independent practices, teaching methods, and educational resources. This enhances students' learning quality and efficiency, promoting their all-round development.

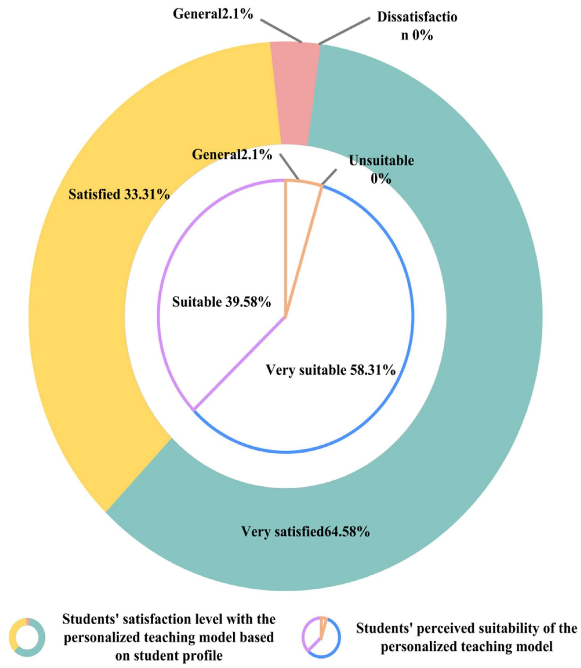

Figure 4. Evaluation Results of the Class Students' Satisfaction and Applicability with the Personalized Teaching Model Based on Student Profiling

In conclusion, the classroom student profiling intelligent education model based on big data algorithms, through classifying students' behavioral data to obtain multi-frequency pattern trees for students with different grades, enables teachers to comprehensively grasp students' learning status and the behavioral differences among students with other grades. At the same time, this model helps students comprehensively understand their own learning behaviors and the gaps between them and students with better grades. Based on this, teachers can tailor personalized teaching to help students improve their learning outcomes. Students can effectively correct their learning behaviour issues during personalized teaching, enhancing learning quality and efficiency and achieving comprehensive development.

5. Conclusions

The application of information technology in education has accelerated educational innovation and reform, leading to significant changes in teaching methods for teachers and learning approaches for students. The classroom student profiling intelligent education model based on big data algorithms constructed in this study effectively discovers the correlation within students' behavioral data, thereby improving the accuracy of predicting student behaviors. Additionally, the classification algorithm based on multi-frequency patterns constructs multi-frequency pattern trees for students with different grades, reflecting the behavioral differences among students. Experimental results indicate that the application of the classroom student profiling intelligent education model based on big data algorithms effectively reflects students' learning behaviors, helping teachers and students gain a comprehensive understanding of the learning situation, improve teaching methods, enhance learning quality, and achieve personalized teaching goals. Furthermore, 90% of the students feel that personalized teaching plans based on student profiling meet their learning needs and effectively improve learning efficiency.

Acknowledgements
The study was supported by research on a team resource sharing mechanism based on big data, a Systematic research project on the construction of teachers' teaching innovation teams in National Vocational Colleges (No. tx20200401).

References

[1] Walkington, C., Bernacki, M. L. (2018). Personalization of instruction: Design dimensions and implications for cognition. *The Journal of Experimental Education, 86*(1), 50-68.

[2] Ai, Y., Li, Li. (2020). Research on the value of inclusive education based on big data in promoting the development of minority education. *Journal of Physics: Conference Series, 1550*(3), 032032 (6pp).

[3] Sato, Y., Lzui, K., Yamada, T., et al. (2019). Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. *Expert Systems with Applications*, 247-261.

[4] Khan, A., Ghosh, S. K. (2018). Data mining-based analysis to explore the effect of teaching on student performance. *Education and Information Technologies, 23*, 1677-1697.

[5] Dong Yu., Jun Hong., Jinhua Zhang., Qingbo Niu. (2018). Multi-objective individualized-instruction teaching-learning-based optimization algorithm. *Applied Soft Computing, 62*, 288-314.

[6] Hicham Moad Safhi., Bouchra Frikh., Brahim Ouhbi. (2019). Assessing reliability of big data knowledge discovery process. *Procedia Computer Science, 148*, 30-36.

[7] Sudip Misra., Anandarup Mukherjee., Arijit Roy. (2018). Knowledge discovery for enabling smart Internet of Things: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1276.

[8] Stover, K., Sparrow, A., Siefert, B. (2017). It ain't hard no more!, Individualizing instruction for struggling readers. *Preventing School Failure: Alternative Education for Children and Youth, 61*(1), 14-27.

[9]Logesh, R., V., Subramaniyaswamy, V.,Vijayakumar, Xiong, Li. (2019). Efficient user profiling based intelligent travel recommender system for individual and group of users. *Mobile Networks and Applications 24(3): 1018-1033*.

[10] Hicham Moad Safhi., Bouchra Frikh., Brahim Ouhbi. (2019). Assessing reliability of big data knowledge discovery process. *Procedia Computer Science, 148*, 30-36.