International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

# A Study on Data Mining Techniques to Improve Students' Performance in Higher Education

**Shilpa K[1], Krishna Prasad K[2]**

[1]Research Scholar, College of Computer Science and Information Science, Srinivas University, Mangalore, India
Orcid-ID: 0000-0003-1356-4860
Email: *shilpaktr[at]gmail.com*

[2]Professor, Institute of Computer Science and Information Science, Srinivas University, Mangalore, Karnataka, India
Orcid-ID: 0000-0001-5282-9038
Email: *krishnaprasadkcci[at]srinivasuniversity.edu.in*

**Abstract:** ***Data mining is a method for extracting information from enormous amounts of data. Data mining, also known as knowledge discovery from data, is the swift and straightforward detection of patterns that hint to knowledge that has been implicitly stored or recorded in big databases, data warehouses, the web, other massive data repositories, or information streams. In this essay, several data mining theories, approaches, tactics, and applications are discussed. The purpose of modern management promotion today is to improve the problem of inaccurate information transmission and increase productivity. The primary goal of contemporary administration is to increase the university potential to develop talent and serve society. The primary focus of this work is on the information management systems that colleges and universities construct utilizing data mining. For many disciplines in higher education, data mining offers useful solutions. Due to the abundance of student data that may be utilized to identify illuminating trends on how students learn, the area of education research is continuously growing. To evaluate student performance and assist them in showcasing the students' accomplishments, educational institutions might use educational data mining. This paper reviews various techniques used as knowledge extractors to tackle specific education challenges from large data sets of higher education institutions to the benefit of all educational stakeholders.***

**Keywords:** Data Mining, Higher Education, Learning Behaviour, Clustering, Student Performance

## 1. Introduction

The growth of information technology has resulted in a large amount of databases and data in many different industries. The raw data is transformed into knowledge that can be applied in a number of academic disciplines. There is now a way to save and maintain the vital data for upcoming decision-making thanks to research in information bases and information technology [1]. The hidden linkages and patterns may not always become apparent after reviewing the data. Data gathering, assembly, analysis, and communication are all steps in the multidimensional process of data mining. Utilizing powerful client/server or parallel processing units, data mining software may analyse the massive amounts of data stored in databases. Universities and colleges will move away from experience-based management and toward scientific or information management based on modern management theory and decision-making science, as well as their application in management within universities and colleges. A number of instructional Information Management Systems (IMS) have been created as a result of the situation, which has been staunchly supported by universities and colleges. But when the number of students managed and the time spent using the teaching management information system grows, a lot of management data based on teaching is produced [2]. The continual rise in college enrolment has made education more adaptable and inclusive. The tension between increasing student enrolment and shrinking teaching resources is generating previously unheard-of management issues at the majority of universities and colleges [3].

Due to this unusual circumstance, the IMS is growing in popularity and respect among professors and students as a result of its ability to facilitate productive collaboration. The prior teaching management was characterized by a concentration on an experience-based management style, an emphasis on the specifics of academic instruction, and, to some extent, a disregard for the parallels between management of general operations and education [4]. It lays an undue emphasis on how correct the college curriculum is while giving insufficient weight to the application of an experience-based management approach. For instance, the IMS has not yet been used by the majority of colleges and universities. This crucial modern management tool contrasts sharply with the many management information systems that other universities and colleges are currently creating for the government and enterprises [5]. Inconsistent institutional recognition of the need of management information systems in university administration continues. When dealing with "mountain" data collection, traditional data analysis methods have some limitations, whether in terms of time or place, such as their incapacity to address the problem and the lack of administrators who are easy to grasp. Effective utilization of these data is necessary to prevent "data disasters" that worsen over time and force college administrators to make "decision disaster" decisions. Because of developments in science, technology, and information technology, the IMS is being used more frequently in higher education [6]. In addition to a significant financial and technological investment, DM is a collaborative endeavour that calls for a variety of specialized knowledge. Repeating this procedure is advised. By performing the same acts repeatedly, the essence of things is persistently sought after, and problem-solving methods are continuously prioritized. During the data subdivision and rearrangement process, the chosen records are divided and added [7]. These data records are selected using a grouping technique for data exploration that looks at numerical statistics, time series visualization, neural network (NN), and decision

tree (DT) models. The evaluation of data understanding, data sampling, data exploration, and data adjustment modelling can all benefit from a detailed interpretation, according to more research [8]. The majority of organizations and colleges currently employ IMSs, which have essentially removed the drawbacks and limitations of an antiquated teaching management system.

## 2. Related Work

The most precise method for assessing useful data in the data warehouse is the DM technique. In order to improve decision-making, DM is used to forecast hidden information through an extraction technique. Based on judgments made by the staff, student performance, and administrative decisions, the usage of DM for instructional activities has risen [9]. The paradigm of data-driven knowledge discovery can be used for DM. DM is a broad field that includes a number of subjects, including statistics, artificial intelligence, information technology, learning, data visualization, and retrieval. The educational system has evolved into one that is more balanced as a result of the improved mining application. The idea of Educational Data Mining (EDM) has quickly developed in the context of various types of educational institutions [10]. Additionally, problems with institutional effectiveness and student achievement have been related to an academic analyst. All elements that directly affect college students are included in the EDM. Table 1 includes the analysis of DM methods used in educational applications.

**Table 1:** Studies Related to Educational Data Mining

| Problem definition | DM technique | Methodology |
|---|---|---|
| Prediction of a student's academic performance [11] | Clustering and classification | NN, Support Vector Machine, DT and Naive Bayes (NB). |
| Academic procrastination prediction by the student [12] | Clustering and classification | Random Forest (RF), k-means, NBTree, decision stump, prism, PART, Iterative Dichotomiser 3 (ID3), ZeroR, open source Java implementation (J48), and One rule (Data mining map) (OneR). |
| Prediction of the student's performance [13] | Regression, classification and clustering | Artificial Neural Network (ANN), DT, and NB. |
| Placement prediction for pupils [14] | Regression, classification and clustering | NB, RF, J48, NN, Random Tree (RT), binomial logistic regression, multiple linear regression, regression tree, recursive partitioning, and conditional Inference tree. |
| Prediction of students' final grades [15] | Multiple Regression Analysis (MRA) | Recurrent neural network |
| Prediction of a student's academic performance [16] | Regression and classification | Fuzzy set rules, lasso linear regression and Collaborative Filtering (CF) |
| Prediction of student performance in educational experiment [17] | Classification | Natural language processing and CART-DT |

## 3. Objectives

a) To study various methods and algorithms related to educational data mining's.
b) To understand various data mining techniques and its applications.

## 4. Methodology

The research methodology used is based on a comprehensive analysis of data mining and its application in higher education. The comments of various national and international conferences were taken into consideration when considering data mining applications in the field of higher education. Thanks to the discussions we had with several academics, groups, colleges that offer higher education, and industry professionals, we were able to identify and present the methodologies, procedures, and applications of data mining in higher education [18]. The main objective of higher education institutions is to provide their students with a top-notch education. Learning how to foretell student enrollment in a specific course, student disengagement from traditional classroom instruction, the use of unfair examination techniques, the detection of abnormal values in student test results, and student performance is one way to guarantee the highest level of quality in the higher education system [19]. The higher education system can be viewed from two perspectives. While the second is a motivator from the outside environment, the first is an internal educational incentive. As higher education institutions compete with their for-profit counterparts, the higher education system demonstrates the external environmental drive. Students must therefore be provided with an appropriate road map and instructed to pursue a certain path in order to get the maximum level of education. The internal educational motive is viewed as a step toward the administration of education's improvement [20].

Higher education institutions must work more rapidly, cheaply, efficiently, and with greater flexibility. "Evaluation, planning, registration, consulting, marketing, and examination" were cited in the research as the six crucial procedures in higher education systems. There are one or more sub-processes in every process. As an example, the important sub-processes of the educational "evaluation" process include "student assessment," "lecturer assessment," "industrial training assessment," "course assessment," and "student registration evaluation." Our suggested effort's main objective is to replace the current educational practices with better ones that offer greater benefits than the accepted practices [21]. A sub process of the "consulting" major educational process is "course selection consulting." By applying some of the classification, clustering, or association technique to the set of student taken various courses data, the characteristic patterns of previous students who took specific elective subjects or courses, as well as the association of courses or elective subjects by various type of students, can be extracted as knowledge and stored in a knowledgebase [22]. The better technique that results from this is "classification or association of students to the most appropriate course and elective subject." The method's output can be used by the academic course planner to create more sophisticated strategies for students' course selection as well as by the faculty consultant to suggest the courses that will

benefit students the most [23]. Table 2 lists some applications for educational data mining.

**Table 2:** Educational Data Mining (EDM) Applications and Methodology used

| EDM Applications | Methodology |
|---|---|
| Predicting Student Performance [24] | Naive Bayes, Bayesian Network, Decision Tree, Rule Based Neural Network, K-Nearest Neighbor (KNN), Multilayer Perception, REPTree, OneR, Iterative Dichotomiser 3 (ID3), Random Forest, PART, Logistic Regression |
| Detecting Undesirable Student behavior's [25] | Social Network Analysis (SNA), Naıve Bayesian, JRip |

To uncover connections and patterns that can aid in decision-making, a lot of data is processed using data mining techniques. Despite the widespread confusion between the phrases "data mining" and "knowledge discovery" in regard to databases, data mining is actually a step in the knowledge discovery process [26]. The algorithms that were found when knowledge was drawn from data are listed below:

a) Cluster Analysis: In order to make a group of objects more similar to one another, a procedure known as clustering is used. The use of clustering in the educational industry can assist institutions in putting specific students into performance-matched classes.
b) Classification and Prediction: Classifications is the practice of looking for patterns that can categorize upcoming data into predetermined categories.
c) Factor Analysis: It is a type of statistical technique used to explain discrepancies between observable, correlated data. Confirmatory Factor Analysis (CFA) and Exploratory Factor Analysis (EFA) are its two subtypes. The underlying structure of variables is revealed through EFA. The variables used in CFA indicators are chosen based on prior theory.
d) Regression Analysis: It contains methods for assessing and modeling a number of variables. The conditional probability of the dependent variable with regard to the independent variables is estimated through regression analysis. Prediction is its main application. It investigates the connections between independent and dependent variables. Ordinary least squares regression and linear regression are the two most used techniques.
e) J48 Algorithm: It is utilized for prediction and classification. As one of the most popular Weka tools that offers a more stable balance between precision, speed, and interpretability of findings, J48 algorithm (based on the C4.5 algorithm from machine learning) was chosen for the classification technique. Data is categorized using a decision tree, and by using this decision tree, we can quickly identify the weaker students. The student's outcome was also predicted using categorization learning.
f) ID3 Algorithm: Iterative Dichotomiser3 is referred to as ID3. A mathematical algorithm of this type is used to create a decision tree. Without going backwards, it will generate the tree from the top down.

## 5. Data Mining Techniques

To extract information from a data source and convert it into a comprehensible format for later use, numerous data mining techniques are used. Table 3 lists numerous data mining techniques along with their objectives. Classification is a method for categorizing things. In these sections, the attributes of the objects are listed or any commonalities between the data elements are highlighted [27]. With the help of this data mining technique, it is possible to categorize and summarize the underlying data more precisely across related attributes or product lines.

**Table 3:** Data Mining Techniques

| Tasks | Supervised | Unsupervised |
|---|---|---|
| Classification [28] | Memory based reasoning, genetic algorithm, C&RT, link analysis, C5.0, ANN | Kohonen nets |
| Estimation [29] | ANN, C&RT | - |
| Segmentation [30] | Market basket, analysis, memory based reasoning, link analysis, rule induction | Cluster detection, K-means, generalized rule induction, APRIORI |
| Description [31] | Rule induction, market basket analysis | Spatial visualization |

The number of students enrolled in college has increased significantly as a result of the ongoing growth in university and college enrolment, significantly adding to the workload of the management staff across all areas of colleges. DM technology is now widely employed in a range of businesses, particularly in college and university teaching management systems, where it aids in understanding fundamental student data, master students' learning styles, and effectively organizing teaching curricula. The manual practices used in traditional management methods are no longer sufficient to handle the demands of the modern workplace [32]. Numerous problems exist with this management strategy, including inefficiency and a lack of secrecy. In addition, a substantial amount of data and files will be created over time, creating difficulties in finding, maintaining, and managing them. Every educational institution must have a University Information Management System (UIMS), and managers and decision-makers rely heavily on its information [33]. The UIMS model should therefore provide users with adequate information and query options. Table 4 gives summary on various data mining techniques.

**Table 4:** Data Mining Techniques Summary

| |
|---|
| Fuzzy logic, Fuzzy clustering, Fuzzy Neural Network (FNN), Neural networks, multilayer perceptron (MLP), Decision Tree, Logistic Regression, Random Forest Classifier, Naïve Bayes Classifier, Support Vector Machine (SVM), K-Nearest-Neighbor (KNN), Associative classification model, Bayesian probabilistic tensor factorization, Latent factor model, Extreme Gradient Boosting (XGBoost), Genetic algorithms, Extra Trees Regressor, Gradient Boosting Regressor, Label Propagation Gaussian NB, Linear Discriminant Analysis, Ada Boost Regressor, Dragonfly Algorithm (DA) for classification models and DTGA-BP (decision-tree-propagation back propagation genetic algorithm) |

There are many uses for DM technology in modern management education, and it is currently the most crucial effort and objective in the advancement of management

education in Indian universities. Before completely engaging teachers in their auxiliary and leading responsibilities, instructors should pay close attention to the physiological and psychological qualities of their students. By hand, some information is incomplete, which can make workers' jobs more challenging and the data more confused. Users should extract data so that it can be applied to people's lives more effectively [34]. Using identification technology and statistical approaches, DM technology extracts the data information required for teaching from a vast volume of geographic data based on a geographical database. The most useful data processing method is ultimately found by continuously analyzing and recognizing these data, offering functional departments of teaching management and leaders of universities and colleges with scientific data support [35]. Mathematical statistics, databases, and artificial intelligence (AI) are three strong technical pillars that have been developed as DM and knowledge discovery research has advanced [36]. The main research areas in data mining and knowledge discovery at the moment include fundamental theory, reuse and maintenance of discovery knowledge, data warehouse, discovery algorithm, visualization technology, quantitative and qualitative exchange model, knowledge representation method, knowledge discovery in unstructured and semi-structured data, and online DM. The term "generalized knowledge" refers to generalized descriptive knowledge of category properties. It determines the information it represents based on the microscopic properties of the data, with universality, higher level concept, medium view, and macro view, which displays the commonality of comparable objects and is the generalization, refinement, and abstraction of data [37]. The concept of interdependencies or connections between events is expressed through association. When two or more qualities are interconnected, it is possible to forecast the value of one quality based on the values of the others [38]. Finding the most well-known association rules involves two steps. Iteratively identifying all frequent item sets is the initial step, and frequent item sets must have a support rate at least equal to the user's lowest value. Building rules from often occurring item sets with a believability equal to or higher than the user's lowest value is the first step. The identification or discovery of all frequent item sets forms the basis of the association rule discovery process [39]. The identification or discovery of all frequent item sets, which also happens to be the portion with the most calculations, forms the core of the association rule discovery technique.

## 6. SWOC Analysis

**Strengths:**

- Predictive modeling, made possible by data mining techniques, enables institutions to forecast student outcomes and quickly spot students who are at danger.
- Using data insights to customize educational experiences enhances student engagement and performance.
- By evaluating data to comprehend course demand, teacher performance, and infrastructure demands, institutions may maximize resource allocation.
- For administrators and educators, data mining offers a powerful system of decision support, assisting with long-term planning and policy creation.

**Weaknesses:**

- Data that is erroneous or incomplete can produce flawed conclusions. Data quality assurance is a never-ending challenge.
- The deployment of data-driven decision-making procedures may be hampered by resistance from professors and personnel.
- It can be difficult to use sensitive student data because of privacy issues and ethical concerns.
- It can be expensive up front to set up the necessary infrastructure and train the necessary employees for data mining.

**Opportunities:**

- The creation of early intervention programs to assist struggling kids and promote academic success is made possible through data mining.
- Curriculum design, instructional strategies, and institutional policies can all benefit from constant data analysis and feedback.
- Data mining and adaptive learning systems can be combined to give students more individualized learning opportunities.
- Research into educational trends, learning patterns, and efficient teaching methods are made possible through data mining.

**Challenges:**

- Sensitive student data must be protected, which a difficult task is given the rise of data breaches and privacy worries.
- Due to interoperability difficulties, integrating data from several systems within a single organization can be difficult.
- The number of employees with the essential abilities to effectively adopt and manage data mining techniques may be in short supply.
- A reductionist view of education can result from a dependence on data to the exclusion of qualitative factors and the human element.

## 7. Finding and Suggestions

The following is a list of data mining methods that can be applied to learning analytics [40-43].

- A student's relationships with other students and teachers in a learning setting can be used to predict how well they will succeed.
- The danger of dropouts can be identified by observing students' behavior, and retention strategies can be implemented before the course even begins.
- The complexity of reports on educational data increases as its volume climbs as well. By quickly scanning the graphic reports produced by data visualization tools, it is feasible to spot patterns and linkages in the data.
- Learning systems have the ability to respond to student input immediately and intelligently, enhancing engagement and performance.
- Based on the interests of the students determined by an analysis of their activity, new courses may be suggested to them. Students won't select foolish occupations that they might not find interesting as a result of this.

- Recognizing the actions of students during sports or community-based activities that help them become student models.
- Collaboration and grouping of students, assessment of students' ability levels, social network analysis, concept mapping, and development of course materials, planning and scheduling, find the success rate for students.
- Assume that a student will gain all the knowledge they require about any course or basic facts about your institution.
- Modules That Are Customized Can Improve Academic Results Through a tailored curriculum, data analytics can improve students' learning and growth.
- Teachers can better understand a student's interests and seek to improve their educational experience by using the knowledge gained from behavior analysis and other evaluations.
- With the right information, it can investigate which type of learning is most effective for your students.
- This knowledge can also enlighten people to strategies for enhancing pupils' general academic experiences.

## 8. Conclusions

A possible analytical technique called Data Mining (DM) can assist educational institutions in managing student results, allocating staff and resources, and establishing alumni programs more efficiently. By identifying hidden patterns in massive databases, community colleges and universities can develop models that precisely predict the behavior of population clusters. By using these prediction models, educational institutions may address problems like transfers and retention, as well as marketing and alumni relations [44]. As educational data generation increases and curriculum design improves, a constructive feedback loop will be established to support the logical alignment of curricular modules. The application of DM theory to university information databases can be enhanced, and the suitable DM algorithms can be adjusted to take into account these databases' characteristics [45]. In order to play a larger role in talent evaluation, discipline echelon construction, post management, and the design of management information policy, DM is conducted from a variety of views, including multilayer and multidimensional.

## References

[1] Femke Nijland, Olga Firssova, Stefan Robbers, Marjan Vermeulen (2023). The networked student. Teaching and Teacher Education, Vol. 131, pp. 1-11.
[2] Schumacher, C., Ifenthaler, D (2018). The importance of students' motivational dispositions for designing learning analytics. Journal of Computing in Higher Education volume, Vol. 30, pp. 599–619.
[3] Selamat, A., Ibrahim, R., & Razak, S. A. (2018). A systematic literature review on data mining applications in higher education. International Journal of Data Mining & Knowledge Management Process, 8(3), pp. 41-59.
[4] Sharma, M., Chugh, R., & Jha, G. K. (2019). Predicting student performance using data mining techniques: A survey. Expert Systems with Applications, Vol. 115, pp. 597-617.
[5] Suthar, P., & Singh, M. (2019). A systematic literature review on data mining techniques for predicting student performance. International Journal of Computer Applications, Vol. 182(47), pp. 15-19.
[6] Thota, M., & Reinhart, R. F. (2017). Data mining applications in higher education. Journal of Education and Learning, Vol. 6(4), pp. 168-180.
[7] Veeramani, R., & Vairavasundaram, S. (2016). Data mining in higher education: A study using educational data mining (EDM) techniques. International Journal of Applied Engineering Research, Vol. 11(6), pp. 4013-4016.
[8] Zhang, J., Zhou, L., & Zhou, N. (2019). Applications of data mining in education: Review and opportunities. Journal of Educational Technology Development and Exchange, Vol. 12(1), pp. 91-104.
[9] Haizhao G. (2020). Research on university educational administration management based on data mining technology. Digital World. Vol. 176(6), pp. 213–3.
[10] Nazarenko MA, Tatiana VK. (2017). Big data in modern higher education. Benefits and criticism. 2017 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS). IEEE. pp. 676–679.
[11] Qian Z, Rui F. (2019). Overview of the application of data mining in the information management of universities and colleges. Data Min. Vol. 9(1), pp. 1–7.
[12] Kaiwei Z. (2018). Research on the application of data mining technology in university educational administration. China Manag Inf. Vol. 21(4), pp. 123–4.
[13] Jach T, Magiera E, Froelich W. (2015). Application of HADOOP to store and process big data gathered from an urban water distribution system. Proc Eng. 2015; Vol. 119, pp. 1375–80.
[14] Liu X, Ma Y, Huang G, Zhao J, Mei H, Liu Y. (2014). Data-driven composition for service-oriented situational web applications. IEEE Trans Serv Comput. Vol. 8(1), pp. 2–16.
[15] Liang J. (2021). Research on education management and decision optimization based on cloud computing from the perspective of big data. In 2021 4th International Conference on Information Systems and Computer Aided Education; 2021. pp. 404–7.
[16] K. Prakash, and J. P. Prasad. (2021). Mobile and Location Based Service using Augmented Reality: A Review", European Journal of Electrical Engineering and Computer Science (EJECE). Vol. 5(2), pp. 13–18.
[17] Garg L, McClean S, Meenan B, Millard P. (2009). Non-homogeneous Markov models for sequential pattern mining of healthcare data. IMA J Manag Mathematics. Vol. 20(4), pp. 327–44.
[18] Haizhao G. Research on university educational administration management based on data mining technology. Digital World. 2020; 176(6):213–3.
[19] Arunachalam AS, Velmurugan T. Analyzing student performance using evolutionary artificial neural network algorithm. Int J Eng & Technol. 2018; 7(2.26):67–73.
[20] Durairaj M, Vijitha C. Educational data mining for prediction of student performance using clustering algorithms. Int J Comput Sci Inf Technol. 2014; 5(4):5987–91.

[21] Akram A, Fu C, Li Y, Javed MY, Lin R, Jiang Y, et al. Predicting students' academic procrastination in blended learning course using homework submission data. IEEE Access. 2019; 7:102487–98.
[22] Sana B, Siddiqui IF, Arain QA. Analyzing students' academic performance through educational data mining. 3c Tecnol glosas innovación Apl a la pyme. 2019; 8(29):402–21.
[23] Bharara S, Sabitha S, Bansal A. Application of learning analytics using clustering data Mining for Students' disposition analysis. Educ Inf Technol. 2018; 23(2):957–84.
[24] Ahmad, Z., & Shahzadi, E. (2018). Prediction of students' academic performance using artificial neural network. Bulletin of Education and Research, 40(3), 157–164.
[25] Alshanqiti, A., & Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. IEEE Access, 8, 203827–203844.
[26] Burgos, C., Campanario, M. L., De, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. Computers and Electrical Engineering, 66(2018), 541–556.
[27] Capuano, N., & Toti, D. (2019). Experimentation of a smart learning system for law based on knowledge discovery and cognitive computing. Computers in Human Behavior, 92, 459–467.
[28] Gök, M. (2017). Predicting academic achievement with machine learning methods. Gazi University Journal of Science Part c: Design and Technology, 5(3), 139–148.
[29] Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using learning analytics to identify successful learners in a blended learning course. International Journal of Technology Enhanced Learning, 5(2), 133–150.
[30] Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A machine-learning approach. Higher Education, 80(5), 875–894.
[31] Ornelas, F., & Ordonez, C. (2017). Predicting student success: A naïve bayesian application to community college data. Technology, Knowledge and Learning, 22(3), 299–315.
[32] Saqr, M., Fors, U., & Tedre, M. (2017). How learning analytics can early predict under-achieving students in a blended medical education course. Medical Teacher, 39(7), 757–767.
[33] Vandamme, J.-P., Meskens, N., & Superby, J.-F. (2007). Predicting academic performance by data mining methods. Education Economics, 15(4), 405–419.
[34] Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. Computers in Human Behavior, 47, 168–181.
[35] Zabriskie, C., Yang, J., DeVore, S., & Stewart, J. (2019). Using machine learning to predict physics course outcomes. Physical Review Physics Education Research, 15(2), 020120.
[36] Tang H, Tan S, and Cheng X (2009). A survey on sentiment detection of reviews. Expert Systems with Applications: An International Journal September 2009, 36(7):10760– 10773.
[37] Abdulsalam, S. (2015). Comparative Analysis of Decision Tree Algorithms for Predicting Undergraduate Students Performance in Computer Programming. Journal of advances in scientific research & its application (JASRA), 2015, Vol. 2, 79-92.
[38] Pandey, M. and S. Taruna, A comparative study of ensemble methods for students' performance modeling. International Journal of Computer Applications, 2014. 103(8).
[39] B. Albreiki, N. Zaki, and H. Alashwal, A systematic literature review of student' performance prediction using machine learning techniques, Educ. Sci., vol. 11, no. 9, p. 552, Sep. 2021.
[40] C. Vieira, Y. Y. Seah, and A. J. Magana, Students' experimentation strategies in design: Is process data enough? Computer Applications in Engineering Education, vol. 26, no. 5, pp. 1903{1914, 2018.
[41] S. Slater, S. Joksimovic, V. Kovanovic, R. S. Baker, and D. Gasevic, Tools for educational data mining: A review, Journal of Educational and Behavioral Statistics, vol. 42, no. 1, pp. 85-106, 2017.
[42] J. Biggs, What the student does: teaching for enhanced learning, Higher Education Research & Development, vol. 31, no. 1, pp. 39-55, 2012.
[43] C. Romero and S. Ventura, Data mining in education, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12-27, 2013.
[44] C. J. Stanny, Reevaluating bloom's taxonomy: What measurable verbs can and cannot say about student learning," Education Sciences, vol. 6, no. 4, p. 37, 2016.
[45] Y. Mao, R. Zhi, F. Khoshnevisan, T. Price, T. Barnes, and M. Chi, One minute is enough: Early prediction of student success and event level difficulty during a novice programming task, The 12th International Conference on Educational Data Mining, pp. 119-128, 2019.