



Strategic Analytics for Predicting Students' Academic Performance Using Cluster Analysis and Bayesian Networks



Shamila Saeedi¹, Darko Božanić², Ramin Safa^{1*}

¹ Department of Computer Engineering, Ayandegan Institute of Higher Education, 4681853617 Tonekabon, Iran

² Military Academy, University of Defence in Belgrade, 11000 Belgrade, Serbia

* Correspondence: Ramin Safa (safa@aihe.ac.ir)

Received: 10-14-2024

Revised: 12-07-2024

Accepted: 12-14-2024

Citation: Saeedi, S., Božanić, D., & Safa, R. (2024). Strategic analytics for predicting students' academic performance using cluster analysis and Bayesian networks. *Educ. Sci. Manag.*, 2(4), 197-214. <https://doi.org/10.56578/esm020402>.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: The evolution of educational systems, marked by an increasing number of institutions, has prompted the integration of advanced data mining techniques to address the limitations of traditional pedagogical models. Predicting students' academic performance, derived from large-scale educational data, has emerged as a critical application within educational data mining (EDM), a multidisciplinary field combining education and computational science. As educational institutions seek to enhance student outcomes and reduce the risk of failure, the ability to anticipate academic performance has gained considerable attention. A novel methodology, employing cluster analysis in combination with Bayesian networks, was introduced to predict student performance and classify academic quality. Students were first categorized into two distinct clusters, followed by the use of Bayesian networks to model and predict academic performance within each cluster. The proposed framework was evaluated against existing approaches using several standard performance metrics, demonstrating its superior accuracy and robustness. This method not only enhances predictive capabilities but also provides a valuable tool for early intervention in educational settings. The results underscore the potential of integrating machine learning techniques with educational data to foster more effective and personalized learning environments.

Keywords: Academic performance prediction; Cluster analysis; Bayesian networks; Educational data mining; Machine learning in education

1. Introduction

The maturity of humans depends on their proper education, and education is a tool to use to reach the highest point of human nobility (Gul & Yucesan, 2022). Looking at the Ministry of Education as the official institution for education for social, political, and cultural development and thinking deeply about it necessitate focusing on the quality of educational services and using state-of-the-art equipment introduced for educational systems (Mahboob et al., 2023). Education quality improvement depends on improving employment status, education, social status, and up-to-date equipment, and tutors need to gain enough knowledge to use these e-learning systems (Shukla et al., 2021). Information and communication technology (ICT) development has created new patterns in education and learning, especially the internet (Rathour et al., 2022). E-learning is a modern educational system in which ICT is utilized for education and learning (Romero & Ventura, 2020). As the main features, e-learning is highly flexible, student-centered, and does not depend on time and location constraints (Gardas & Navimipour, 2022). Providing and establishing human, technological, administrative, social, cultural, managerial, and economic infrastructures are obvious actions to take to start e-learning courses successfully and, ultimately, realize the virtual university concept (Gonçalves et al., 2023). The main challenges and obstacles e-learning faces include cultural, economic, legal, educational, strategic, and technical obstacles, untrue beliefs, content, non-allocation of sufficient budget, lack of internet access for most people, and non-tendency to acquire information and electronic literacy skills (Salloum et al., 2020).

Families, especially those with few kids, are mainly concerned about their academic status and future. One of

the biggest problems of the Ministry of Education and some families is students' failure in education (Wanke et al., 2022). Hence, the Ministry of Education and families should look for a solution to predict students' academic performance (Yağcı, 2022). Records of students exist in schools, including personal information of students and their families, personal features, course feedback, and a sample of exam papers. Students' report cards in the records of schools under study are collected after the exams once the students finish elementary school, getting ready to register for the next educational stage (Amjad et al., 2022). Among these students, those with better academic status qualify for gifted schools. However, a number of students are neither qualified nor have the resources to register in their respective schools despite having excellent report cards.

Failure in education can cause moral disorders in students, making them lose self-esteem and feel stress, and the feeling of stress causes the students to become aggressive or introverted (Adhikari et al., 2022). Schools hold records of each student, which contain information including personal features, course feedback, and a sample of exam papers. This study uses data mining and cluster analysis to predict students' academic performance, and tries to predict the problems a student can face in a course and the possibility of failure in education in the future if he/she keeps studying in the same manner using the information on his/her records (Mahboob et al., 2023; Yağcı, 2022).

Failure in education is one of the primary problems that some families and kids face, given that it gives rise to moral disorders in students (Bago, 2022). For example, students lose self-esteem and feel stressed, which can cause aggressive or introverted behaviors. This study aims to predict students' academic performance using data mining and records specific to each student, expecting to predict their failure in education before it happens and to keep families posed to prevent it.

This study is organized as follows: In section 2, leading papers on fields related to the subject of this research were reviewed. Section 3 gives an overview of initial research theories, concepts, and other necessary items to get better acquainted. In section 4, the presented method was explained in detail, along with a completely clarified flowchart, and the recommended algorithm was explained. In section 5, the introduced model was compared in various dimensions to existing methods using various metrics. Ultimately, the final section concludes this study with a conclusion and discussion.

2. Research Background

The learner model is defined as a representation of beliefs of a computer system about the learner; hence, it is an abstract representation of the learner (Rostaminezhad et al., 2013). Learner modeling consists of a set of views and attitudes that a learner can possess. However, in practice, the learner and user models should be discriminated, which is a more comprehensive model. It is obvious that learners (students) are essential components of a smart educational system.

The user model has a more general form than the learner model, and conducted research can focus on more general behaviors that do not address a specific aspect of learner behaviors. Hence, the user model consists of aspects not limited to teaching (education). Investigating the learner model through an educational plan and learning theory is better. Three main theories of learning include behaviorism, cognitive constructivism, and constructivism. The oldest theory of learning is behaviorism which looks at the learner as a black box. The instructor of behaviorism is like a machine that responds in the face of stimuli. Learning occurs when learners are provoked to respond to a specific stimulus. Hence, learning happens when a learner faces those stimuli repeatedly; his/her correct responses are boosted through rewards, and wrong responses are rejected through punishments.

Based on behaviorism, cognitive constructivism theory assumes that learning consists of acquiring cognitive structures through saving information and processing them (Mosharraf et al., 2017). In other words, learning is defined as reshaping and recorrecting mental representations of intended aspects. Therefore, in cognitive constructivism, an individual member of the learners' group is not defined as a black box; rather, his/her mental representations are defined through cognitive models.

The third theory is constructivism. Whereas the previous two theories are concrete theories of learning in which pre-determined behaviors possessing cognitive structures are transferred to the learner, the third theory is a speculative theory in which learners reconstruct the truth based on acquired experiences. Instead of being transferred, new knowledge is shaped according to previous experiences. Existing mental structures, as well as the learner's beliefs, are employed to interpret events and objects. Each learner is, therefore, expected to construct his/her own reality. But where do traditional Intelligent Tutoring Systems (ITSs) containing a learner model fit into this psychological framework, and how do they relate to one another? A theory in which a test is used to assess a learner's skill level belongs to the theory of behaviorism and the behaviorist viewpoint. Likewise, ITSs that attempt to model a learner's internal state are, in fact, cognitive constructivist. Constructivist theories, however, are not compatible with traditional ITSs. Suppose each learner constructs a reality based on his/her own specific prior experiences and knowledge. Assuming that a pre-determined model can reasonably define such a learner is meaningless. Therefore, this study represents a proposed learner model, which utilizes Bayesian networks in the form of behaviorist and cognitive constructivist theories, given that current ITSs support these two.

White & King (2020) proposed a model for identifying key factors in academic guidance of students using decision tree and neural network algorithms in data mining, and used the model to help facilitate the students' academic guidance and advancements and increase their chances for success. Zaki et al. (2020) investigated the possibilities for education quality improvement in e-learning systems using EDM. Their research mainly aims to utilize data mining to obtain experiences that go beyond those of experts and to use these experiences for academic guidance in e-learning systems. The research also deals with hidden patterns in students' course unit selection and prediction of their grades. In addition, it investigates the effects of activeness, the circumstances and time of entrance, season, etc., in an e-learning management system. Zhang et al. (2019) identified the key factors behind academic slumps using association rules and cluster analysis. The research attempts to implement predictive data mining models to predict students' academic performance based on their personal and academic information. The statistical results in this research, produced from the implementation of models for predicting student status, can be used to discover the most effective factors that cause academic slumps and to help prevent them, as well as to improve the quality of communication between administrators/parents and students, and to improve the quality of education overall. Asif et al. (2017) proposed a method for student academic guidance based on mixed-technique data mining, and used student academic history in guidance school and first-year high school to point them toward appropriate academic majors. Various techniques were used to construct the intended models, such as improved decision trees and nearest-neighbor algorithms. A genetic algorithm was also used to process information gathered from 969 students besides the crisp methodology in MATLAB and clementine.

Several studies have examined and categorized the most important and beloved data mining techniques for improving education and creating personalized education based on data from traditional and distance education systems in recent years, including web-based courses, educational material management systems, and web-based intelligent/adaptive education systems.

Dutt et al. (2017) presented a decision support system based on a multilayered perceptron neural network to help facilitate selecting an appropriate guidance strategy. In the next stage, an evolutionary algorithm was used to validate the knowledge produced by the neural network and evaluate the effectiveness of the specified guidance strategy. The neural network must be built using the fewest layers to prevent system decision-making mistakes (Sorourkhah et al., 2019). The proposed method mitigates many problems and complexities with neural networks by constructing trees out of the information and data. Slater et al. (2017) combined multiple classification algorithms to categorize students and predict their grades based on features extracted from their status. A genetic algorithm was utilized to weigh data features. And as the results point out, the rates for classification and predicting the students' status have improved in this article. Devasia et al. (2016) used data mining techniques like association rule mining and inter-session and intra-session frequent pattern mining to extract useful patterns for instructors, administrators, as well as web managers who evaluate the students' online course activity. A computer-based method was proposed for handling problems in student learning regarding courses in the sciences and for providing students with counseling.

3. Theoretical Foundations and Literature Review

Effective and efficient living in the 21st century requires recognizing the characteristics of this century (Qiu et al., 2023). The main characteristics of this era are the Information Age and the information-oriented community (Guzzo et al., 2023). In this community, information and its management and transformation to base knowledge constitute the foundation of the communities' economy (Muniz, 2022). Such characteristics significantly impact social and economic institutions, based on which social institutions are forced to be reconstructed (Saberhoseini et al., 2022). Institutions for education and learning in general and at higher levels are one of the social institutions that will go through significant changes (Heilporn et al., 2022). Currently, the industry-oriented community is the institution's foundation for education and learning. Graduates of traditional educational systems cannot possess the needed proficiency in an information-oriented community. In the past, people have been educated commensurate with the agricultural and industrial ages. However, such a procedure is not acceptable today. Today, information technology allows people to educate commensurate with needs, considering that it eliminates past constraints, provides us with authority, and allows students to realize their academic needs for learning in a proper time (Zhang et al., 2020). A new approach is needed for education and learning so that students can possess the necessary proficiency in an information-oriented community. Information technology and available tools are needed to implement the new approach (Chansamut, 2021). Such opportunity has been provided given the development of the information network, such as the internet. Using these opportunities quickly and in time can help people progress and develop. E-learning is one of these opportunities. Education planning should be done to help use this opportunity in the best way possible.

3.1 Data Mining

In the last two decades, humans have become more capable of producing and quickly collecting data (Imani et

al., 2022). The following factors have a significant role in these changes: using barcodes for business productions, using computers in business, science, and public services, and developing data collection instruments such as image and text scanners and remote sensing satellites (Roiger, 2017). Data mining can be considered a natural evolutionary process of information technology resulting from an evolutionary process in the dataset industry (Moterased et al., 2021). Data mining uses several scientific fields simultaneously, just like data collection, data management, and data analysis (Leskovec et al., 2020), including dataset technology, artificial intelligence (AI), machine learning, neural networks, statistics, pattern recognition, knowledge-based systems, acquisition of knowledge, information retrieval (IR), high-speed calculations, and data visualization.

Algorithm architectures define the instructions clearly to express the functions. In simple terms, an algorithm is a step-by-step calculation method and machine learning algorithms are used for various types of predication. They are categorized as supervised, unsupervised, semi-supervised, and reinforcement learning includes Artificial Neural Networks (ANNs), decision trees, Bayesian networks, k-nearest neighbors (KNN), and Support Vector Machines (SVMs) (Alsariera et al., 2022; Sharifi et al., 2022).

Bayes algorithm is based on Bayes' theorem and is utilized for making real-time predictions and ensuring that the task is done risk-free (Wang et al., 2021). For example, each school intends to introduce several teams to the Paya Scientific League so that students can participate in the scientific league competition. In this study, students with the potential to attain acceptable scores and ranks were requested to participate in the competition, seeking to save time and increase efficiency. Individuals who are not eager to participate in the competition or are weak were exempted. Students' information was stored in the system, including their grades, the financial status of their families, and their guarantee of presence at the completion. Using this algorithm, students can be informed and invited only to participate in this competition.

3.2 Cluster Analysis

Clustering is the same task as daily categorization people do daily. For example, people place items in the same group due to their similarity and in another group due to their difference. In addition, the same items can belong to various groups based on their model, size, or use (Ebeling et al., 2019). This algorithm can place one item in two or more groups, given the similarities and differences. There is an essential difference between chain clustering and cluster analysis. In cluster analysis, the clusters are formed based on similarity. However, in chain clustering, clusters are formed based on the model. In chain clustering, each step is connected to the next one, just like a chain, meaning that the next step starts upon passing the first step. For example, in the case of Instagram, people should first register on Instagram and enter their personal information. Then after becoming an Instagram user to visit pages, Instagram shows pages they are interested in. As a result, friends with the same conditions can be shown (Shen et al., 2016).

3.3 Evaluation Metrics

There are various evaluation metrics to measure and evaluate classification systems. These metrics include classification accuracy, recall, precision, and error rate. Before introducing these metrics, it is better to get familiarized with concepts, including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which are utilized in the metrics.

TP is a percentage of the number of members in class X that the classification system correctly classifies as members of class X. TN is a percentage of the number of members in other classes classified correctly as not belonging to class X. FP is a percentage of the number of members in other classes classified incorrectly as members of class X. FN is a percentage of the number of members in class X classified incorrectly as members of other classes. Positive (P) is the total number of class members classified correctly. Negative (N) is the total number of class members classified incorrectly.

The evaluation metric recall is the accuracy of a classification system in correctly classifying members of class X, which are correctly classified as a member of class X, and is calculated as follows:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (1)$$

The evaluation metric precision is the percentage of members classified as class X members that truly belong to class X and is calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Additionally, classification accuracy is another metric to evaluate the classification systems' performance, owning an expansive and comprehensive view and domain of the performance of classification systems. It includes all the members who are classified correctly. Classification accuracy is calculated using the following equation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

In addition to those said above, F-score is another metric that is the weighted mean between accuracy and recall metrics. It is used to determine the efficiency of classification systems. It is calculated as follows:

$$F - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Ultimately, the error rate of the suggested composition is calculated using the following formula:

$$\text{Error rate} = 100 - \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \quad (5)$$

4. Methodology

There has been a limited number of studies until today focusing on predicting students' academic performance using their records and report cards in school. Considering the potential of data mining methods in predicting future values, this study tries to use these technologies and cluster analysis to predict students' failure in education and determine effective features. Ultimately, useful recommendations can be provided to parents through data analysis.

The suggested method in this research consists of the following parts to predict students' academic performance:

- Clustering students
- Predicting students' academic performance using a new algorithm

The main goal of the suggested method is first to find similar users using cluster analysis and determine the academic status of students given the users' academic status in the same cluster.

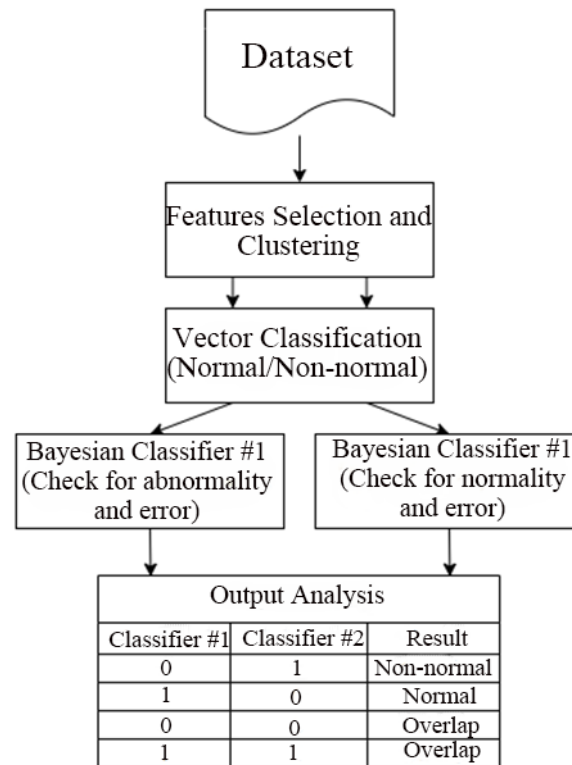


Figure 1. Proposed diagram

The suggested flowchart can be observed in the following Figure 1. Table 1 illustrates students' general features and information. This form is available for all elementary school students under investigation and in the records. Additionally, each $F1$, $F2$, $F3$, $F4$, and $F5$ is a student's feature.

Table 1. Students' general features and information

No.	F1			F2			F3			F4			F5			
	Students being interested in a course			Families checking on how their kids are educated			Family Status			Absence			Parents' Education Stage			
	Very Good	Good	Average	Average	Good		Divorced Parents	Conflicted Family	Good	Dead Parents	Excused	Unexcused	Good	Average	Good	Average
													Bachelor's degree and lower (Father)	Master's degree and higher (father)	Bachelor's degree and lower (mother)	Master's degree and higher (mother)
Student 1	*				*				*		*		*			*
Student 2	*				*	*			*	*		*	*			*
Student 3	*			*			*		*	*		*	*			*
...																
Student n	*				*				*		*		*		*	*

It can be noted that the school under investigation is a public school; therefore, parents either have a high income or average income, and low-income families do not exist. An intelligence quotient (IQ) test was conducted to register students in said school. It was found that students are either highly intelligent or have average intelligence, and students with low IQs do not exist. Students living with conflicted families have parents who disagree with each other, or the kids are only enrolled in public school because their parents do not bear the responsibility for their kids' education. The result of academic status can be categorized as follows: a) Students with very good academic status enroll in a gifted school. b) Students with very good report cards but lower academic status than the first group enroll in a middle-range school for their next grade. c) Students with much lower academic status than other said groups enroll in a low-range school for their next grade.

4.1 Clustering Similar Students

Cluster analysis was used in this study so that data can benefit from a better structure and coherence. This method was also used to predict students' academic performance with high accuracy, given that the dataset consists of students and there are similarities between students. Usually, similar students have the same academic performance. The first purpose of this study is to cluster the students and make predictions based on students of the same cluster, benefiting from a high accuracy. Other algorithms, such as machine learning and association rule learning, can also be beneficial in this context. However, cluster analysis helps the data set to have better similarities for intended students.

The k-means clustering algorithm has a complexity of $O(I \times K \times n)$, where I is the number of iterations, K is the number of clusters, and n is the number of samples. K-means clustering is a non-hierarchical, flat, and algorithmic method that starts its search in a local environment, dividing data into each K cluster with a specific feature. This way, data in each cluster can be similar to each other to a feasible extent, and the difference between data of different clusters can have the highest value.

Each cluster has a centroid consisting of data, and all data are placed at the least distance from the cluster centroid. K-means clustering is an iterative clustering algorithm that minimizes the total distance between objects inside the cluster and the centroid. Next, the objects are placed in various clusters until the distance between objects does not change. The result of this clustering algorithm is completely separated clusters that do not resemble each other under no circumstances. For example, it's assumed that students with the features mentioned in Table 2 exist for academic status. This is a hypothetical table, aiming to elaborate on the suggested method step-by-step using an example.

The following figures are given for each said feature: $F1$ (students being interested in a course): 18 to 20 is considered for the "very good" section, 15 to 17 for the "good" section, and 12 to 14 for the "average" section. $F2$ (families checking on how their kids are educated): 15 to 17 is considered for the "average" section, and 18 to 20 for the "good" section. $F3$: Regarding family status, 2 is considered for "divorced families," 3 for "conflicted families," 10 for "good families," and 4 for "dead parents." Regarding students' absence, 10 is considered for "students without absence," -1 for "excused absences," and -2 for "unexcused absences." $F4$: Regarding financial status, 3 is considered for the "average" section, 6 for the "good" section, 6 to 10 for students with "good IQ"; and 3 to 5 for students with "average IQ." $F5$: According to the laws of the Ministry of Education, each educational stage has 2 points. In other words, a diploma has 2 points, an associate degree 4 points, a bachelor's degree 6

points, a master's degree 8 points, and a Ph.D. 10 points. Academic degrees of parents, bachelor's degree and lower, are placed in one section in Table 1.

Table 2. Data of several students with similar features

Student Code	F1	F2	F3	F4	F5
Student 1	20	20	12	20	12
Student 2	17	18	4	15	12
Student 3	16	19	8	20	12
Student 4	19	17	9	16	20
Student 5	15	20	8	20	20
Student 6	18	18	2	17	10
Student 7	19	16	20	17	12
Student 8	14	20	9	20	12
Student 9	17	19	1	15	14
Student 10	19	17	8	16	14

Next, student clustering was conducted below. First, a mean vector was constructed based on input data. The mean value of each feature was calculated by dividing the total values of each feature by the number of features. For example, the mean value of F5 in Table 2 is calculated as follows:

$$F5 = (12+12+12+20+20+10+12+12+14+14)/10=17.4$$

Three cluster heads were selected based on this vector, including a vector with the greatest positive distance from the mean vector, the least distance from the mean vector, and the greatest negative distance from the mean vector. The distance of all vectors from the mean vector was calculated using the following formula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

$$d(x, y) = \sqrt{((17.4 - 20)^2 + (18.9 - 20)^2 + (7.9 - 12)^2 + (17.6 - 20)^2 + (14.8 - 12)^2)} = 6.19758$$

The similarity between vectors increased as the distance decreased. For example, the distance of student 1 from the mean vector was calculated as follows based on Table 2.

The distances of all data in Table 2 from the mean vector were calculated and are illustrated in Table 3.

Table 3. Distance from the mean value

Student Code	Distance
Student 1	6.19758
Student 2	5.547972
Student 3	3.94715
Student 4	6.1229011
Student 5	6.57114
Student 6	10.73175
Student 7	12.569
Student 8	5.98163
Student 9	7.360706
Student 10	3.062678

The average distance is 5.889. Hence, vector data of students 7, 8, and 10 were selected as cluster heads considering: a) Student 7 has the greatest positive distance compared to the mean value of 5.889 (the greatest figure among figures larger than 5.889). b) Student 10 has the greatest negative distance compared to the mean value of 5.889 (the smallest figure among figures smaller than 5.889). c) Student 8 has the least distance from the mean vector.

Table 4. Results of clustering data from Table 2

Members	Cluster Heads
Student 6	Student 7
Students 1, 2, 4, 5, and 9	Student 8
Student 3	Student 10
Student 10	3.062678

The distance of each data point from selected cluster heads was calculated using the said formula to select members of each cluster. Clusters absorbed vectors with the least distance, and the results are elaborated in Table 4.

The distance between each vector was calculated from cluster heads for the next input data. For a lower distance between the vector and average distances, the vector was absorbed by one of the cluster heads; otherwise, it was selected as one of the new cluster heads.

4.2 Prediction Using New Algorithms

In the previous sections, students were clustered, given their features. This section aims to predict students' academic performance using the Bayesian networks.

Bayesian network is a learning process based on statistical learning theory, which is one of the best machine learning approaches in data mining. This method has been successful in various tasks, such as data classification, pattern recognition, content classification, face recognition on images, recognition of figures written by hand, and bioinformatics. In fact, the Bayesian network is a binary classifier that separates two classes using a linear boundary. This method uses all bands and an optimization algorithm to obtain samples that form the boundary of classes. These samples are called support vectors. A number of learning points with the least distance from the decision boundary can be considered a subset to define decision boundaries and as a support vector. Assume that two data classes have a total of $x_i, i=1, \dots, L$ learning points (x_i is a vector). These two classes are tagged with $y=\pm 1$. The optimal margin classifier calculates the decision boundary of two completely separated classes. In this method, the linear boundary between two classes is calculated so that all samples of the +1 class are on one side of the boundary, and samples of the -1 class are on the other side of the boundary. The decision boundary should be selected so that the distance of the nearest educational samples from each other in each class, orthogonal to the decision boundary, is maximized to a feasible extent. A linear decision boundary can be defined as follows in general: $w \cdot x + b = 0$. x is a point on the decision boundary, and w is an n -dimensional vector orthogonal to the decision boundary. $b/|w|$ is the distance between the origin and decision boundary, and $w \cdot x$ represents the inner product of two vectors of w and x .

The equality is still established by multiplying both sides of the equation by a constant. Finding the nearest educational samples of two classes is the first step to calculating optimal decision boundaries. Next, the distance between these points, orthogonal to the boundaries that separate two classes completely, is calculated. The optimal decision boundary has the maximum margin. The optimal decision boundary is calculated by solving the following optimization problem.

$$\min_{w,b} \min_{i=1,\dots,L} \left[y_i \frac{(w \cdot x_i + b)}{|w|} \right] \quad (7)$$

The above-written equation can be written as follows using a set of mathematical operations:

$$\min_{w,b} \frac{1}{2} |w|^2, y_i (w \cdot x + b) - 1 \geq 0, i=1,\dots,L \quad (8)$$

It is difficult to reach a solution to the above-written optimization problem. A Lagrange multiplier was utilized to write this optimization problem in the form of the following equation aiming at simplifying it. λ_i are the coefficients of the Lagrange multiplier.

$$\max_{\lambda_1, \dots, \lambda_L} \left[-\frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L \lambda_i y_i (x_i \cdot x_j) y_j \lambda_j + \sum_{i=1}^L \lambda_i \right] \quad (9)$$

$i=1,\dots,L$

$$w = \sum_{i=1}^L \lambda_i y_i = 0 \quad (10)$$

w can be calculated using the following equation after solving the above-written problem and finding the coefficients of the Lagrange multiplier.

$$w = \sum_{i=1}^L \lambda_i y_i x_i \quad (11)$$

λ_i is larger than zero for support vectors and equal to zero for other points. Therefore, considering the above equation and λ_i being zero for x_i other than support vectors, a limited number of educational points, which are the same as support vectors, are needed to obtain the decision boundary. Not all the points are needed.

In such circumstances, the following operations were conducted:

- a) A cluster for the new student was specified.
- b) Students in a cluster were separated as normal and non-normal. Students who did not fail in education were considered normal; otherwise, non-normal. Next, data obtained in this section was considered a training set.
- c) New students and students in the same cluster were evaluated using the suggested algorithm to make predictions as follows:

- Firstly, training data were specified.
- Training data were separated as normal and not-normal.
- Two classifiers were utilized to determine the academic status of each student.

The first classifier specifies whether the new student fails in education. The output of this classifier is either one or zero. In other words, the output “1” means that failure in education occurs, and the output “0” means undetermined. In this classifier, data from students who fail in education were used in training data.

• The second classifier specifies whether the new student has a good academic status. The output of this classifier is either one or zero. In other words, the output “1” means that failure in education does not occur, and the output “0” means undetermined. This classifier uses data from students with good academic status in the training set.

- After finding the output of each classifier, the final results were studied as follows:

If the first classifier gains output “1” and the second classifier gains output “0,” the student definitely fails in education.

If the first classifier gains output “0” and the second classifier gains output “1,” the student does not definitely fail in education.

If both classifiers gain either output “1” or “0,” an indefinite circumstance occurs. To eliminate such circumstances, the vector distance of the student from the average vector of normal and non-normal data was calculated. If any of them have a lesser value, the student is placed in that class. Next, the suggested algorithm was explained using an example.

The suggested algorithm uses the training data set to execute this task. Several students were in this data set, and whether they fail in education was determined. Students with better academic status have class “1,” and those with inferior academic status have class “0,” which are depicted in the following table.

Table 5 illustrates 15 education vectors as an example. In this table, each row represents a student with $F1$, $F2$, $F3$, ..., features.

Table 6 illustrates test data, showing the students’ current status. However, their normal or non-normal status is not specified.

This study aims to specify new students’ academic status using the suggested algorithm. Two columns were added to Table 5 as the output of classifier No. 1 and classifier No. 2 to determine the current vectors’ status. Using the output of each classifier, outputs of the current vector status were determined. In the naive Bayes classifier, all shared features were first summed and then divided by the number of features. Then, the data test was compared. Table 7 shows the average features of students in each class.

Table 5. Train set sample (Users in the same cluster)

Class	$F1$	$F2$	$F3$	$F4$	$F5$
0	18	16	14	15	8
1	18	18	17	13	8
0	16	16	13	14	20
1	19	19	19	18	16
1	20	18	17	15	12
1	19	19	20	13	12
0	13	15	2	13	10
1	20	18	20	15	4
0	13	15	0	12	2
1	15	16	1	10	12
1	13	15	2	10	20
1	18	20	20	11	14
0	12	15	14	11	6
1	18	17	20	15	4
0	17	17	20	20	6

Table 6. Test data sample

Class	F1	F2	F3	F4	F5
---	17	15	2	13	2
---	12	17	0	10	20
---	16	19	17	14	12
---	19	18	14	18	8
---	18	20	13	20	4
---	20	19	20	15	4

$$F1(class1) = (18 + 19 + 20 + 19 + 20 + 13 + 18 + 17)/8 = 18/6$$

$$F1(class0) = (18 + 16 + 13 + 13 + 15 + 13 + 12)/7 = 14/2$$

Table 7. The average features of students in each class

Class	F1	F2	F3	F4	F5
1	18.6	18.2	19.1	14.8	8.7
0	14.2	15.4	6.5	12.1	11.14

4.3 Classifier No. 1

The negative selection classification algorithm (NSCA) with a variable length and real values and detection systems with a variable radius were used to design classifier No. 1. The radius of self-samples (RS) is the main parameter that significantly affects classification efficiency, and it is a crucial element in learning capability (classifier generalization). It also plays a significant role in both the NSCA and the positive selection classification algorithm (PSCA) with real values. If this classifier generates a positive output, it most likely shows the high quality of educational services, and a suitable warning should be issued. RS was calculated below for all test data. The similarity ratio of each test data to data from the education section was calculated in non-normal circumstances.

The similarity ratio is the number of shared features of both the vector and training data vectors. It is illustrated in percentage at the end, and the highest value is considered the similarity ratio. RS of a vector is the average similarity ratio of all vectors. This classifier generates an output equal to 1 when the similarity ratio of a vector exceeds RS.

Table 8 shows the similarity ratio for test data to non-normal training data. It can be noted that the evaluation of each data feature was tested using Table 5 to achieve the ratio of RS similarity to normal and non-normal data. This way, the similarity ratio to normal and non-normal data was obtained. Additionally, each of these features was considered as 20%.

Table 8. The ratio of test data similarity to non-normal data

Ratio of Similarity to Non-Normal Data	F1	F2	F3	F4	F5
60%	17	15	2	13	2
80%	12	17	0	10	20
20%	16	19	17	14	12
0%	19	18	14	18	8
20%	18	20	13	20	4
0%	20	19	20	15	4

Given the similarity ratio of test data in Table 7, obtained based on non-normal data from training data, RS is as follows:

$$RS = (0.6 + 0.8 + 0.2 + 0 + 0.2 + 0)/6 \text{ RS} = 30\%$$

Based on what was said previously, classifier No. 1 generated output “1” for vectors with a ratio of similarity larger than RS, as illustrated in Table 8.

4.4 Classifier No. 2

The second classifier examines whether the current vectors of the students are in a normal state or not. This classifier operates based on the normal training data. If the vector is normal, it outputs one; otherwise, it outputs zero. Table 9 shows the percentage of similarity between the test data and the normal training data.

Table 9. The ratio of test data similarity to normal data

Ratio of Similarity to Normal Data	F1	F2	F3	F4	F5
60%	17	15	2	13	2
20%	12	17	0	10	20
80%	16	19	17	14	12
100%	19	18	14	18	8
40%	18	20	13	20	4
100%	20	19	20	15	4

Based on the percentage of similarity in the table for the test data, obtained based on the normal training data, the value of RS is equal to:

$$RS = (0.6 + 0.2 + 0.8 + 1 + 0.4 + 1)/6 \quad RS = 66\%$$

Therefore, vectors that have a similarity value greater than RS received one output from the second classifier. Table 10 shows the outputs generated by classifiers.

Table 10. Outputs generated by classifiers

Classifier #1	Classifier #2	F1	F2	F3	F4	F5
1	1	2	13	2	15	17
1	0	20	10	0	17	12
0	1	12	14	17	19	16
0	1	8	18	14	18	19
0	0	4	20	13	20	18
0	1	4	15	20	19	20

4.5 Mechanism to Select the Best Classifier

During the test phase, four modes occurred for a new input sample: Classifier No. 1, which covers a non-self area, generated the output “1,” and classifier No. 2, which covers the self area, generated the output “0.” In this case, it can be 100% said that the new sample belongs to class 2, is a malformed sample, and generates a warning system.

Classifier No. 1 generates the output “0,” and classifier No. 2 generates the output “1.” In this case, it can be 100% said that the new sample belongs to class 1 and is a normal sample. Both classifiers generate either output “0” or “1”; such circumstances are called indefinite. A method was used for selecting one of the classifiers explained below. Table 11 shows the result of the initial evaluation of classifiers.

Table 11. Result of the initial evaluation

Class	Classifier #1	Classifier #2	F1	F2	F3	F4	F5
Indefinite	1	1	17	15	2	13	2
Non-normal	1	0	12	17	0	10	20
Normal	0	1	16	19	17	14	12
Normal	0	1	19	18	14	18	8
Indefinite	0	0	18	20	13	20	4
Normal	0	1	20	19	20	15	4

Vectors with normal or non-normal classes have a definite status. However, the methods said were used below for indefinite circumstances.

4.6 Indefinite Circumstances

This occurs when both classifiers generate either output “1” or “0.” A survey was used in such circumstances. If a group has the highest number of mature data resembling test data, its tag was selected for test data. However, suppose the similarity ratio was the same for both normal and non-normal groups. In that case, if the vector has only one feature different from the education section vector, it can be called similar.

As seen in Table 11, some vectors are overlapped. The said method needs to be used to specify whether they are normal or non-normal; this matter was explained vector by vector below. In Table 12, vectors in overlapping mode are depicted along with the similarity ratio.

Based on Table 13, the status of one vector was specified. However, one vector still remained indefinite. The said method was used to calculate the new similarity ratio in such a case. In this method, if the vector has only one

feature different from the education section vector, it can be called similar.

Table 12. Vector status in overlapping mode

Ratio of Similarity to Normal Data	Ratio of Similarity to Non-Normal Data	Class	F1	F2	F3	F4	F5
60%	60%	Indefinite	17	15	2	13	2
40%	20%	Normal	18	20	13	20	4

Table 13. Similarity ratio

Class	Output of Classifier No. 1	Output of Classifier No. 2	F1	F2	F3	F4	F5
Normal	20%	40%	17	15	2	13	2

5. Results and Comparison with Available Methods

This section evaluates results acquired in the simulation using several metrics. MATLAB was used for simulation tasks, and several phases were considered for each evaluation metric. The results were implemented for various data and compared with available methods. It is worth noting that the data evaluated in each phase is larger than the dataset. Simulation was used several times to obtain results in each phase, and the average-acquired results were considered the output result.

Chuan et al. (2017) used decision trees and neural networks to introduce a model for identifying factors affecting students' performance by studying their records and consulting them. A model was suggested with the aid of decision tree algorithms and neural networks in data mining that helps students' performance and increases their success. In the study by Francis & Babu (2019), data mining techniques, such as association rule mining, intersession, and intra-session frequent pattern mining, were utilized to extract useful patterns for tutors, heads of education, and web managers who evaluate students' online activities. A computer-based method was suggested to eliminate students' learning problems in scientific courses and consult them.

5.1 Recall

Recall is one of the important metrics to evaluate extracted rules. To calculate this metric, the output of average values is depicted for each phase in Figure 2.

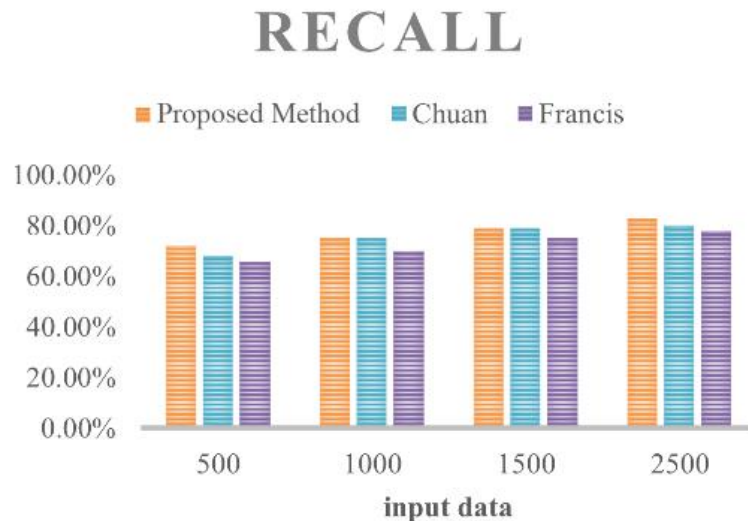


Figure 2. Comparison of recall metrics

5.2 Precision

Figure 3 depicts the comparison result of the above-said metric. In each phase, the average of the above-said metric values was calculated. The extracted rule benefits from a higher assurance as the value increases.

The result shows that the suggested method for the above-said metric performs better than the previous methods.

PRECISION

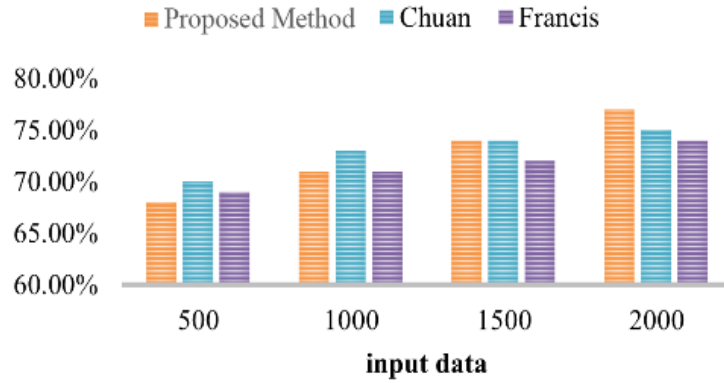


Figure 3. Comparison of precision metrics

5.3 F-Measure

This metric was calculated using the following equation given two metrics of recall and precision:

$$F - \text{Measure} = \frac{2 * \text{Reacall} * \text{Presicion}}{\text{Reacall} * \text{Presicion}} \quad (12)$$

Figure 4 depicts the metric results for simulation and comparison.

F1-MEASURE

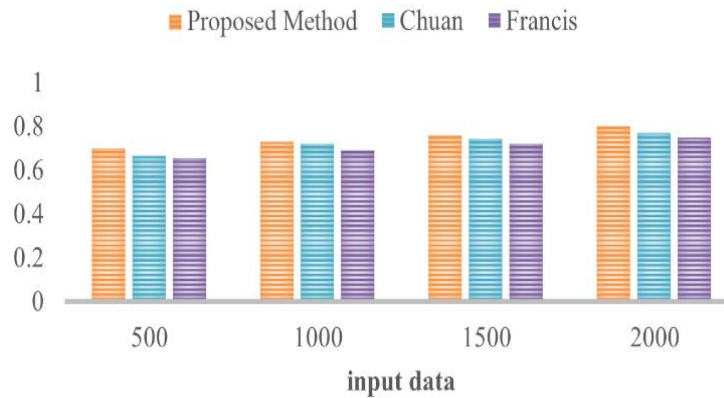


Figure 4. Comparison of the F1-measure metric

The results of the simulation, shown in the figure, mean that the suggested method is improved by 12%.

5.4 FPR

FPR is one of the significantly important evaluation metrics, and it shows the error rate of the intended method in determining non-correct modes. In other words, this metric shows the error rate for modes that are supposed to be determined as wrong, but the method could not recognize that. The lesser the metric, the better the result. This metric is calculated as follows: $FPR = FP/N$, where N is the total number of non-normal vectors, and FP is the number of data recognized wrongly as positive. The simulation was conducted in four phases to evaluate this metric, and the number of data investigated in each phase increased. Figure 5 depicts the results.

The simulation result shows that the suggested method performed better, and Figure 6 depicts the resulting average for all simulation modes.

FALSE POSITIVE RATE

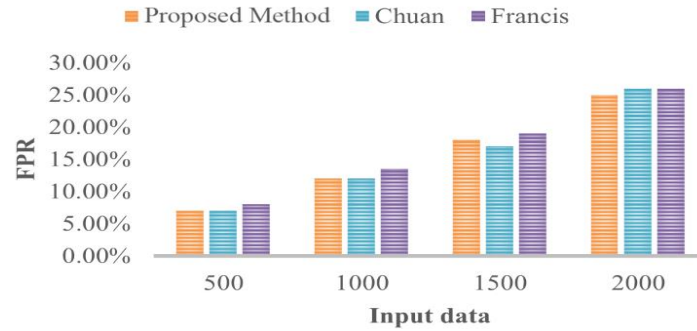


Figure 5. FPR metric

FALSE POSITIVE RATE

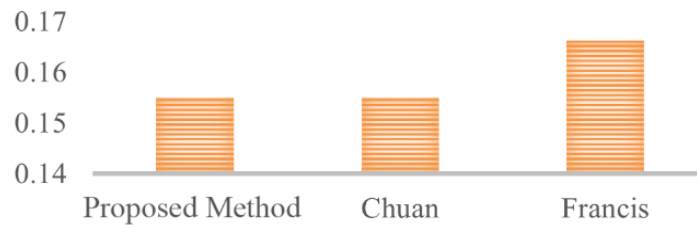


Figure 6. The mean of FPR

5.5 FRR

This metric is utilized to investigate how wrong the suggested method is in determining normal circumstances, incorrectly reporting non-risky modes as unsafe. This metric shows what percentage of correct modes are incorrectly recognized as incorrect by the recognition system, issuing a wrong warning. This metric is calculated using the following equation: $FRR = FN/P$, where P is the total number of positive data, and FN is the number of data to be recognized as negative incorrectly. The lesser the metric, the better the result.

FALSE REJECT RATE

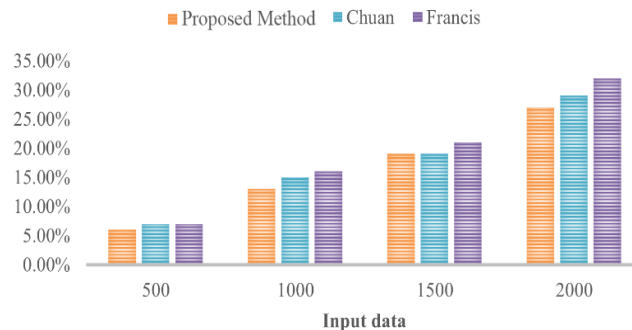


Figure 7. Comparison result of the FRR metric

Like the previous metric, the simulation was conducted in four phases, and Figure 7 depicts the results.

Simulation shows that the suggested method in this metric improved less compared to the two previous methods. Figure 8 depicts the average of the metric.

FALSE REJECT RATE

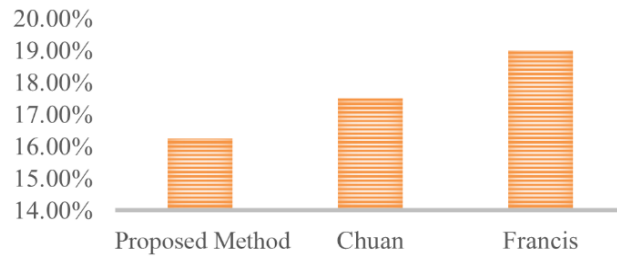


Figure 8. The mean of FRR

5.6 Accuracy

Classification accuracy is another metric to evaluate classification systems' performance, providing a more extensive and comprehensive view of their performance. It is defined as the number of correct classifications. Classification accuracy is calculated using the following equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

The simulation result in the metric shows that the suggested method performed better compared to the previous two methods, and Figure 9 depicts the mean accuracy of each method.

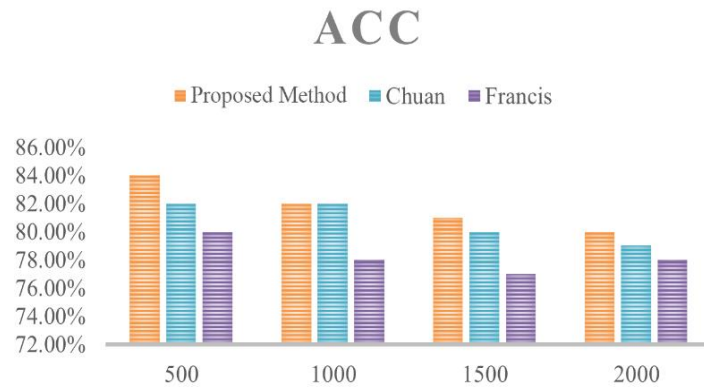


Figure 9. Comparison result of the accuracy metric

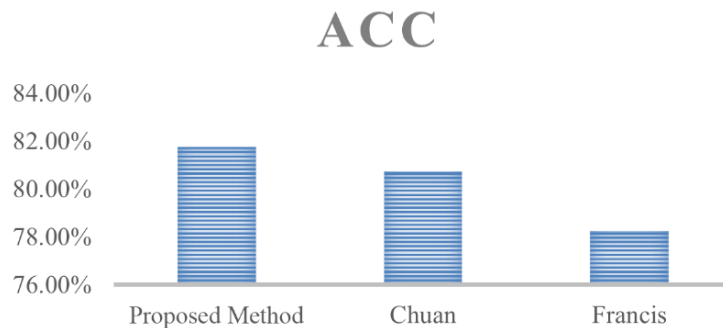


Figure 10. The mean accuracy

Simulation is one of the crucial parts of scientific research and can be used to prove the performance of the suggested methods. In this section, the method suggested in Section 4 was simulated using MATLAB, analyzed, and compared with the other two available methods based on several important metrics. Simulation results using

10-fold cross-validation and the comparison show that the proposed method performed better in analyzing and predicting behavior and can be used in related settings. Figure 10 shows the mean accuracy.

6. Conclusion

In this study, a new way was proposed to suggest educational services, consisting of users' feature selection, clustering, and classification. In the first part, fruitful data were utilized for feature selection that possessed value in most records. The presented method was used to investigate which feature is found more frequently statistically to place the final choice. Feature selection helps the suggested algorithm work with valid input data, increasing output accuracy, which is one reason the suggested method is improved during simulations. The clustering technique places identical users in a group to get the aid of similarity to make predictions and increase output accuracy. The main algorithm was based on Bayesian networks, in which a new view of the operation of Bayesian networks was formed.

In other words, the user can receive the rules by entering the value of the intended parameters at the least time possible. The results show that the proposed model can be used as a service-suggesting system in the educational domain and also has benefits in terms of accuracy and speed compared to other methods.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Adhikari, A., Bhattacharyya, S., Basu, S., & Bhattacharya, R. (2022). Evaluating the performance of primary schools in India: Evidence from West Bengal. *Int. J. Productivity Perform. Manage.*, 71(7), 2630-2658. <https://doi.org/10.1108/IJPPM-07-2020-0374>.
- Alsariera, Y. A., Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Comput. Intell. Neurosci.*, 2022(1), 4151487. <https://doi.org/10.1155/2022/4151487>.
- Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., & Gani, A. (2022). Data mining techniques to analyze the impact of social media on academic performance of high school students. *Wirel. Commun. Mob. Comput.*, 2022(1), 9299115. <https://doi.org/10.1155/2022/9299115>.
- Asif, R., Mercer, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Comput. Educ.*, 113, 177-194. <https://doi.org/10.1016/j.compedu.2017.05.007>.
- Bago, B. A. (2022). Effect of single parenthood in students' academic performance; A case of selected secondary schools in Bitereko Sub County Mitooma District. *IAA J. Social Sci. (IAA-JSS)*, 8(1), 216-226.
- Chansamut, A. (2021). Information system model for educational management in supply chain for Thai higher education institutions. *Int. J. Res. Ind. Eng.*, 10(2), 87-94. <https://doi.org/10.22105/RIEJ.2021.285518.1225>.
- Chuan, Y. Y., Husain, W., & Shahiri, A. M. (2017). An exploratory study on students' performance classification using hybrid of decision tree and naïve Bayes approaches. In *Advances in Information and Communication Technology: Proceedings of the International Conference, ICTA 2016* (pp. 142-152). Lausanne: Springer International Publishing.
- Devasia, T., Vinushree, T. P., & Hegde, V. (2016). Prediction of students performance using educational data mining. In *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)* (pp. 91-95). Ernakulam, India.
- Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991-16005. <https://doi.org/10.1109/ACCESS.2017.2654247>.
- Ebeling, H., Atek, H., Edge, A. C., Kaiser, N., Kneib, J. P. R., Limousin, M., McPartland, C., Repp, A., Richard, J. P., & Toft, S. (2019). Beyond MACS: A snapshot survey of the most massive clusters of galaxies at $z=0.5-1$. *HST Proposal*, 15843.
- Francis, B. K. & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *J. Med. Syst.*, 43(6), 162. <https://doi.org/10.1007/s10916-019-1295-4>.
- Gardas, B. B. & Navimipour, N. J. (2022). Performance evaluation of higher education system amid COVID-19: A threat or an opportunity? *Kybernetes*, 51(8), 2508-2528. <https://doi.org/10.1108/K-10-2020-0713>.
- Gonçalves, M. J. A., Tavares, C., Terra, A. L., Moreira da Silva, M., Bernardes, Ó., Valente, I., & Lopes, I. C. (2023). Digital tools and methods to enhance learning: The digitools project. In *Perspectives and Trends in*

- Education and Technology: Selected Papers from ICITED 2022* (pp. 399-413). Singapore: Springer Nature Singapore.
- Gul, M. & Yucesan, M. (2022). Performance evaluation of Turkish Universities by an integrated Bayesian BWM-TOPSIS model. *Socio-Econ. Plann. Sci.*, 80, 101173. <https://doi.org/10.1016/j.seps.2021.101173>.
- Guzzo, T., Caschera, M. C., Ferri, F., & Grifoni, P. (2023). Analysis of the digital educational scenario in Italian high schools during the pandemic: Challenges and emerging tools. *Sustainability*, 15(2), 1426. <https://doi.org/10.3390/su15021426>.
- Heilporn, G., Lakhali, S., & B  lisle, M. (2022). Examining effects of instructional strategies on student engagement in blended online courses. *J. Comput. Assisted Learn.*, 38(6), 1657-1673. <https://doi.org/10.1111/jcal.12701>.
- Imani, A., Abbasi, M., Ahang, F., Ghaffari, H., & Mehdi, M. (2022). Customer segmentation to identify key customers based on RFM model by using data mining techniques. *Int. J. Res. Ind. Eng.*, 11(1), 62-76. <https://doi.org/10.22105/riej.2021.291738.1229>.
- Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). *Mining of Massive Data Sets*. Cambridge, UK, Cambridge University Press.
- Mahboob, K., Asif, R., & Haider, N. G. (2023). Quality enhancement at higher education institutions by early identifying students at risk using data mining. *Mehran Univ. Res. J. Eng. Technol.*, 42(1), 120-136. <https://doi.org/10.22581/muet1982.2301.12>.
- Mosharraf, M., Taghiyareh, F., & Alaei, S. (2017). Investigating elearning research trends in Iran via automatic semantic network generation. *J. Global Inf. Technol. Manage.*, 20(2), 91-109. <https://doi.org/10.1080/1097198X.2017.1321355>.
- Moterased, M., Sajadi, S. M., Davari, A., & Zali, M. R. (2021). Toward prediction of entrepreneurial exit in Iran: A study based on GEM 2008-2019 data and approach of machine learning algorithms. *Big Data Comput. Visions*, 1(3), 111-127. <https://doi.org/10.22105/bdcv.2021.142089>.
- Muniz, S. M. (2022). Deployment of agriculture 4.0 with the integration of IoT. *Comput. Algorithms Numer. Dimensions*, 1(3), 122-125. <https://doi.org/10.22105/cand.2022.161803>.
- Qiu, P., Sorourkhah, A., Kausar, N., Cagin, T., & Edalatpanah, S. A. (2023). Simplifying the complexity in the problem of choosing the best private-sector partner. *Systems*, 11(2), 80. <https://doi.org/10.3390/systems11020080>.
- Rathour, L., Obradovic, D., Tiwari, S. K., Mishra, L. N., & Mishra, V. N. (2022). Visualization method in mathematics classes. *Comput. Algorithms Numer. Dimensions*, 1(4), 141-146. <https://doi.org/10.22105/cand.2022.159701>.
- Roiger, R. J. (2017). *Data Mining: A Tutorial-Based Primer*. Boca Raton, US, Chapman and Hall/CRC.
- Romero, C. & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplin. Rev. Data Min. Knowl. Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>.
- Rostaminezhad, M. A., Mozayani, N., Norozi, D., & Iziy, M. (2013). Factors related to e-learner dropout: Case study of IUST elearning center. *Procedia-Social Behav. Sci.*, 83, 522-527. <https://doi.org/10.1016/j.sbspro.2013.06.100>.
- Saberhoseini, S. F., Edalatpanah, S. A., & Sorourkhah, A. (2022). Choosing the best private-sector partner according to the risk factors in neutrosophic environment. *Big Data Comput. Visions*, 2(2), 61-68. <https://doi.org/10.22105/bdcv.2022.334005.1075>.
- Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020). Mining in educational data: Review and future directions. In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)* (pp. 92-102). Switzerland: Springer International Publishing.
- Sharifi, A. M., Khalili Damghani, K., Abdi, F., & Sardar, S. (2022). A hybrid model for predicting bitcoin price using machine learning and metaheuristic algorithms. *J. Appl. Res. Ind. Eng.*, 9(1), 134-150. <https://doi.org/10.22105/jarie.2021.291175.1343>.
- Shen, J., Hao, X., Liang, Z., Liu, Y., Wang, W., & Shao, L. (2016). Real-time superpixel segmentation by DBSCAN clustering algorithm. *IEEE Trans. Image Process.*, 25(12), 5933-5942. <https://doi.org/10.1109/TIP.2016.2616302>.
- Shukla, R., Khalilian, B., & Partouvi, S. (2021). Academic progress monitoring through neural network. *Big Data Comput. Visions*, 1(1), 1-6. <https://doi.org/10.22105/bdcv.2021.142228>.
- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *J. Educational Behav. Stat.*, 42(1), 85-106. <https://doi.org/10.3102/1076998616666808>.
- Sorourkhah, A., Babaie-Kafaki, S., Azar, A., & Nikabadi, M. S. (2019). A fuzzy-weighted approach to the problem of selecting the right strategy using the robustness analysis (Case study: Iran automotive industry). *Fuzzy Inf. Eng.*, 11(1), 39-53. <https://doi.org/10.1080/16168658.2021.1886811>.
- Wang, Y., Xiao, Z., Tiong, R. L., & Zhang, L. (2021). Data-driven quantification of public-private partnership experience levels under uncertainty with Bayesian hierarchical model. *Appl. Soft Comput.*, 103, 107176. <https://doi.org/10.1016/j.asoc.2021.107176>.
- Wanke, P. F., Antunes, J. J., Miano, V. Y., Couto, C. L. D., & Mixon, F. G. (2022). Measuring higher education

- performance in Brazil: Government indicators of performance vs ideal solution efficiency measures. *Int. J. Productivity Perform. Manage.*, 71(6), 2479-2495. <https://doi.org/10.1108/IJPPM-11-2020-0602>.
- White, E. & King, L. (2020). Shaping scholarly communication guidance channels to meet the research needs and skills of doctoral students at Kwame Nkrumah University of Science and Technology. *J. Academic Librarianship*, 46(1), 102081. <https://doi.org/10.1016/j.acalib.2019.102081>.
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learn. Env.*, 9(1), 11. <https://doi.org/10.1186/s40561-022-00192-z>.
- Zaki, M. J., Meira Jr, W., & Meira, W. (2020). *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge, UK, Cambridge University Press.
- Zhang, G., Wu, J., & Zhu, Q. (2020). Performance evaluation and enrollment quota allocation for higher education institutions in China. *Eval. Program Plann.*, 81, 101821. <https://doi.org/10.1016/j.evalprogplan.2020.101821>.
- Zhang, M., Zhu, J., Wang, Z., & Chen, Y. (2019). Providing personalized learning guidance in MOOCs by multi-source data analysis. *World Wide Web*, 22, 1189-1219. <https://doi.org/10.1007/s11280-018-0559-0>.