

Research Article

Sohaib Latif*[#], Fang XianWen[#], and Li-li Wang

Intelligent decision support system approach for predicting the performance of students based on three-level machine learning technique

<https://doi.org/10.1515/jisys-2020-0065>

received June 23, 2020; accepted March 19, 2021

Abstract: In this research work, a user-friendly decision support framework is developed to analyze the behavior of Pakistani students in academics. The purpose of this article is to analyze the performance of the Pakistani students using an intelligent decision support system (DSS) based on the three-level machine learning (ML) technique. The neural network used a three-level classifier approach for the prediction of Pakistani student achievement. A self-recorded dataset of 1,011 respondents of graduate students of English and Physics courses are used. The ten interviews along with ten questions were conducted to determine the perception of the individual student. The chi-squared (χ) test was applied to test statistical significance of the questionnaire. The statistical calculations and computation of data were performed by using the statistical package of IBM SPSS version 21.0. The seven different algorithms were tested to improve the data classification. The Java-based environment was used for the development of numerous prediction classifiers. C4.5 algorithm shows the finest accuracy, whereas Naïve Bayes (NB) algorithm shows the least. The results depict that the classifier's efficiency was improved by using a three-level proposed scheme from 83.2% to 88.8%. This prediction has shown remarkable results when compared with the individual level classifier technique of ML. This improvement in the accuracy of DSSs is used to identify more efficiently the gray areas in the education stratum of Pakistan. This will pave a path for making policies in the higher education system of Pakistan. The presented framework can be deployed on different platforms under numerous operating systems.

Keywords: neural network, real-world data mining, decision tree, machine learning

1 Introduction

Any country in this world is characterized and recognized based on its education standards and quality. According to the World Bank survey, the population growth rate is rapidly increasing in Asia day by day. Pakistan has the highest population growth rate in the world [4], surpassing India and China during the last

[#] These authors contributed equally to this work.

* **Corresponding author: Sohaib Latif**, School of Mathematics and Big Data, Department of Computer Science, Anhui University of Science and Technology, Huainan, Anhui 232001, China, e-mail: sohaib.latif@aust.edu.cn

Fang XianWen: School of Mathematics and Big Data, Department of Computer Science, Anhui University of Science and Technology, Huainan, Anhui 232001, China, e-mail: xwfang@aust.edu.cn

Li-li Wang: The Key Laboratory of Embedded System and Service Computing, Ministry of Education (Tongji University), Shanghai 201804, China, e-mail: wanglili198301@163.com

decade. The consequence of an increase in population growth has also increased the demand for the quality of education. For the development of any country, there is a dire need to improve the standards and protocols of their education and research systems and methods. One possible solution to this problem is to increase the education sector without any standards. Whenever we look at this option, an alarming situation will occur. However, this creates an emergency in the education field. In addition to this, the education data mining (EDM) research field has been growing exponentially and gaining popularity nowadays and the sketch of this modern pattern is given in Figure 1. Because it has the capacity for enhancing the standard of education and training institutions, educationalists and scientists are trying to incorporate the factors from sophisticated and complex data of the education sector. More specifically, the traditional database only caters to abstract level queries. To improve the educational curriculum design, we need to predict the true behavior of student performance. It also helps out the plan interventions for academic support and student regarding the curriculum. Data mining is an important technique for the analysis of datasets and transforms the information structure into understanding form. Different computational techniques are used to predict a student's behavior.

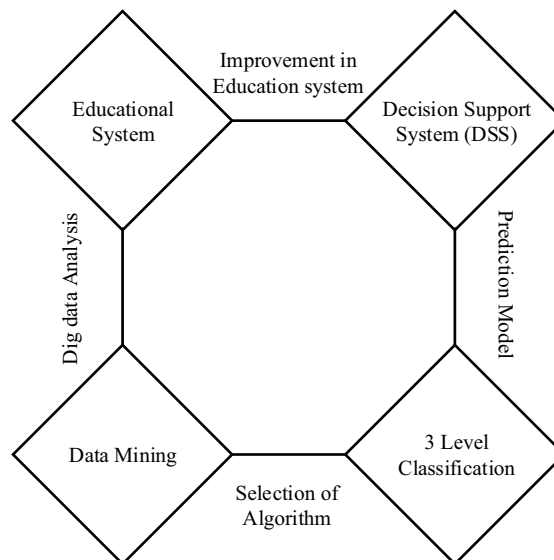


Figure 1: The major steps involved to predict the student performance framework.

The significance of EDM has established on the way that it permits teachers and specialists to separate valuable ends from refined and confounded inquiries, such as “find the students who are at risk in failing the examinations” or “find the students who will exhibit excellent performance” in which conventional dataset queries were not tackled [25].

The EDM is one of the most essential methods where intelligent steps are implemented for the extraction of data patterns in student's databases for the discovery of key features. This new exploration field has developed exponentially and picked up prominence in the advanced instructive period in view of its capability to improve the nature of instructive foundations and frameworks.

In the most recent decade, research concentrated on creating proficient and accurate decision support systems (DSS) for foreseeing the understudy's future scholastic exhibition [10,15,24,31,33,39]. All the more scientifically, a scholarly DSS is an information-based data framework to catch, deal with, and break down data which influences or is proposed to influence dynamic performed by individuals in the extent of an expert assignment designated by a client [7]. The advancement of a scholarly DSS is noteworthy to understudies, instructors, and instructive associations, and it will be more significant if information mined from the understudy's presentation is accessible for instructive chiefs in their dynamic cycle.

1. Intelligent decision support system (IDSS) is used to measure the performance of students.
2. The real-world data are analyzed and collected from different undergraduate students of Pakistani University.
3. In this article, seven different machine learning (ML) classifier algorithms (Back-Propagation [BP], sequential minimal optimization [SMO], Naïve Bayes [NB], C4.5, RIPPER (JRip), 3-NN, and Voting) are used to predict the performance of students.
4. The three-level classifier provides the better performance as compared to the two-level classifier.
5. It is user friendly software, which predicts the performance of the students
6. Evaluate and analyze the various teaching methods on the performance of students by forecasting the results using a simulation of the proposed model.

Our essential objective is to help the scholastic errand of effectively foreseeing the understudies execution in the last assessments of the school year. Besides, leaders can assess different instructive procedures and produce estimates by using a few information. The remainder of the article is organized as follows: Section 2 presents a short related writing audit of the proposed work. Section 3, depicts the three-level classifier AI strategy utilized in this article. Inevitably, Section 4 discusses about the key highlights of our choice help program, and the discoveries of the proposed research are given in Section 6.

2 Literature review

During the most recent decade, the use of information digging procedures for the improvement of exact and effective DSS has given helpful results and results that help with tending to numerous issues and issues in the instructive space. Romero and Ventura [35,36] and Baker and Yacef [6] have given some broad audits of various kinds of instructive frameworks and how information mining can be effectively applied to every one of them. All the more explicitly, they depicted in detail the way toward mining learning information, just as how to apply the information mining methods, for example, insights, perception, grouping, bunching, and affiliation rule mining. As of late Dutt et al. [12] introduced a survey of how EDM tries to find new experiences into learning with new apparatuses and procedures, so those understanding the effect the movement of specialists in all degrees of training. In Table 1, we quickly present the absolute most agent techniques for Applied EDM and Learning diagnostic that depend on an exhaustive writing audit.

Table 1: A brief comparison of some related papers is made here on the relevant database and existing techniques used for the similar type of work

S. no.	Paper title	Data-set size	Algorithms	Test options	Algorithm higher accuracy
1	Performance predication of engineering students using decision trees [20]	346	J48	Cross validation	J48
2	Data mining applications: a comparative study for predicting student's performance [40]	48	ID3, C4.5, CART	Cross validation	CART
3	A promising classification method for predicting distance student's performance [14]	194	J48 Rtree JRip, OneR, NB	—	—

Higher education institutions of Pakistan overwhelmed with a large amount of data of students, containing the information of student enrollment, the number of courses introduced in the educational institutions, student achievement in each of the courses, and annual result statistics [30]. The growing use of computerized systems for record-keeping brings a new term, "big-data" [2]. The extensively high amount of interrelated data about any

entity that can be used for analytical analysis is called big data [28]. It is difficult to perform an analytical process on big data for making decisions about curricula reforms and restructuring the education system.

The implementation of intelligent methods is essential for extracting data patterns and analytical information to discover hidden knowledge from student databases [9]. This new research area has become popular and grown exponentially in the new era of modern education, the reason behind its potential and capacity building for familiarizing improvement in the quality of education systems [3].

Conventional education systems are different from the newly emerged ML tools assisted systems where the availability of multi-dimensional data is in abundance [22]. IDSSs are widely used in various computer science applications for intelligent decision making [26]. The application of a DSS is mainly concentrated on improving the learning process by the development of accurate models that predict student's characteristics and performance [13]. The importance of this system is founded on the fact that it allows educators and researchers to extract useful conclusions from sophisticated and complicated questions such as "find the students who 'will' exhibit poor performance" in which traditional database queries cannot be applied [34].

The secondary education system in Pakistan is a two-tier system in which the first 2 years cover the general topics and introduction of the majors of degree [18]. The remaining 2 years contain focused learning on the selected subjects, and this period is called higher education system [29]. So the last 2 years of higher secondary education have immense significance and decisive factors for the life of any student; it acts as a connecting bridge from school learning to higher education, provided by different universities and higher educational institutes [38]. Therefore, the capacity to monitor the student's academic performance and achievement is considered a highly important factor for the identification of possible bad performance that could lead to the decay of education performance [37]. During the last years, the researchers aim to develop an efficient and precise decision support system to how the projected student's academic performance [11]. More analytically, an academic DSS is a knowledge-based information system to capture, handle and analyze information that affects or is intended to affect decision making performed by people in the scope of a professional task appointed by a user [27]. The development and implementation of an academic decision support system are significant to students, educators, and educational organizations [21]. Management of education must continue to implement and evaluate over an ongoing basis to improve the quality of institutions [5]. This study aims to implement the ML tools and framework for DSS to improve the education system of Pakistan that provides help in the performance evaluation student's in the final examinations.

Livieris *et al.* [22] implement software for predicting the student's performance. They reported the course of "Mathematics" for the prediction of the newly enrolled student of Lyceum. They performed the experiments on the variety of classification algorithms, which depicts that the neural network classifier exhibited the best accuracy and consistent behavior. Along this line, in ref. [23] the authors predicted student's performance using user-friendly DSS software. A hybrid prediction approach is used to predict a student's performance. In this work, four different ML algorithms are utilized as a simple voting scheme. Their experimental results show that the DM can predict student progress and performance deeply.

Bunching calculations are likewise applied to voluminous information sizes, for example, large information. The idea of large information alludes to voluminous, gigantic amounts of information both in computerized and physical organizations that can be put away in incidental stores, for example, records of understudy's tests or assessments just as accounting records by Dutt *et al* [12]. An informational collection whose computational size surpasses the handling furthest reaches of the product can be ordered as large information as proposed by Abaker *et al.* [17]. A few examinations have been led in the past that gives the point-by-point experiences into the utilization of customary information mining calculations like bunching, forecast, and relationship to tame the sheer voluminous intensity of large information by Hormigo [18]. Comprehensively, the instructive framework can be delegated two sorts; block or mortar-based customary homerooms and computerized virtual study halls otherwise called Learning Management Systems, electronic versatile hypermedia frameworks [8], and smart coaching frameworks [1].

3 Methodology

The point of this investigation is to build up a three-level ML procedure dependent on a choice emotionally supportive network for anticipating the understudy's exhibition at the last assessments. For this reason, we have received the strategy pipeline which comprises three phases. The primary phase of the proposed strategy worries with information assortment and information readiness. The subsequent stage presents the three-level arrangement conspire. In the last stage, we assess the classifier execution of our proposed three-level arrangement calculation with the most mainstream and habitually used calculations by directing a progression of analyses. The proposed strategy work process is clarified in Figure 2.

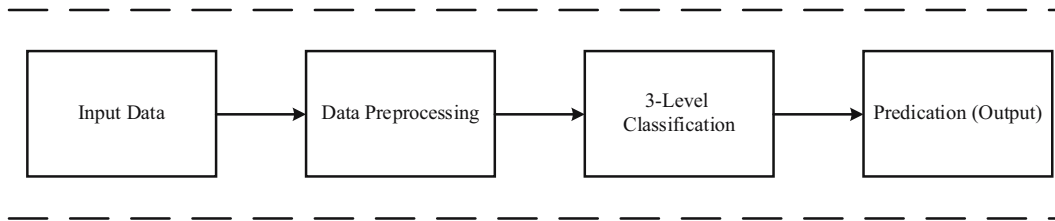


Figure 2: The general pipeline block used in this method.

3.1 Data preprocessing

The preprocessing step comprises the data acquisition. The raw form of information is collected from different graduate students. For this purpose, the different questionnaires are conducted. The quality of the questionnaire is tested by applying the statistical chi-squared (χ) test [32]. After processing the raw information, we manage it in the dataset form and its details are given in Section 3.1.1.

3.1.1 Dataset

The study used a set of data collected from 1,011 respondents in graduate-level students having a result of “English” and “Physics” courses. Table 2 enlists the set of parameters used throughout the analysis, which relate to evaluation details of the participants such as marks obtained in percentage, attendance, class participation, number of students failing, and cumulative comprehension of the curriculum. During each course, the students are assessed by verbal communication and written examination. This investigation lasts for 3 h. The main features that are considered during the data are as follows:

1. The ten interviews with ten questions were included to determine the perception of every student from the program.
2. The 15-min tests are given to solve the verbal questions and thorough problem solving.

The 3-h tests lead to multiple theoretical and advanced mathematical problems requiring methods to be solved and evaluated critically. Finally, each student's overall 5-semester grade discusses the student's interest and its success.

The student attribute values were categorized into four equal segment schemes (i.e., Fail (0–3), Good (4–5), Very good (6–8), and Excellent (9–10)). On the basis of these values, the student's performance is measured.

Nevertheless, since it is significant to an instructor to apprehend poor students within the midst of the instructional cycle, dataset has been created in this context primarily based on the parameters furnished in Table 3 and the elegance allocation.

- $DATA_A$: This includes the characteristics related to the success of the first semester graduates.
- $DATA_{AB}$: This includes the characteristics related to the first and second semester results of the graduates.

Table 2: Activities performed in the dataset and about videos counts

S. no.	Attributes	Type	Value
(i) First semester student data			
1	The number obtained in percentage in first semester	Real	[0–9]
2	First-semester attendance	Real	[0–9]
3	Class participation in first semester	Real	[0–9]
4	Number of the student who fails	Real	[0–9]
5	Curriculum understanding	Real	[0–9]
(ii) Second semester student data			
1	The number obtained in percentage in second semester	Real	[0–9]
2	Attendance in second semester	Real	[0–9]
3	Class participation in second semester	Real	[0–9]
4	Number of the student who fails	Real	[0–9]
5	Curriculum understanding	Real	[0–9]

3.2 Proposed model

The new three-level classification scheme has a great challenge to implement with three different stages. Three-level identification strategies are deterministic ML devices intended to provide precision than one-level tools at the cost of some classification scheme complications. The study uses the proposed classification scheme on Level-A to classify the students either they are pass or fail. The Level-B classifier determines the reason for the failure. This stage also determines the curriculum design challenges. The Level-C stage determines the performance of the student as shown in Figure 3. This performance was judged in the entire class for a particular semester. Furthermore, this algorithm explicitly determines whether the student performs between 0–3 (Fail), 4–5 (Good), 6–8 (Very Good), and 9–10 (Excellent). To assess the passing and failing classifier used for student A-Levels. In the case, final examinations have declared Fail at the A-Level classifier then B and C classifiers are the best option to determine the most likely reason for failure. It is worth bearing in mind that the A-Level classifier's judgment is coarser and broader, while the B-Level and C-Level's (i.e., level 3) decision is finer and portrays the explanation for an ineffective student so that either the problem was with the interpretation of the program, the person's disinterest of the teaching method.

4 Experimental setup

This section aims to describe the results of the simulations and experiments performed for the proposed scheme, and it provides a detailed performance analysis of the proposed scheme for three-level-based DSS

Table 3: A fragment from a dataset, containing Cp index and Cu index

S. no.	Roll no.	Name	Total mark	Cp index	Cu index
1	100	Arfan	68	4	3
2	102	Nadia	46	3	2
3	103	Hamza	88	6	7
4	104	Jamal	73	8	8
5	105	Khazala	89	7	9
6	106	Yasmeen	34	2	3
7	107	Amad	78	6	7
8	108	Rashid	69	6	6

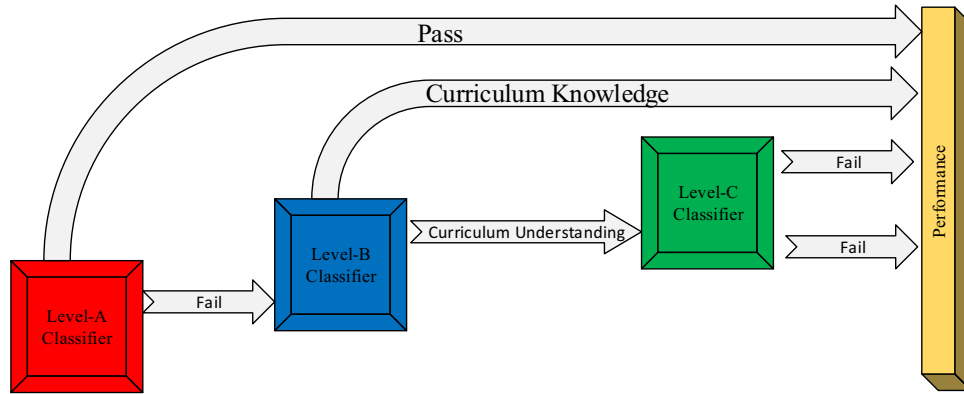


Figure 3: This diagram is basically the representation of hole proposed pipeline of our approach.

to predict the student’s performance. For this purpose, we have divided the experimental setup into the following stages that are given as follows.

- Simulation environment
- Performance matrices
- Quantitative result.

4.1 Simulation environment

For this purpose, JAVA-based prototype is used for the implementation. WEKA ML Toolkit is also used, which is compatible with Java Virtual Machine. The framework of this prototype is given in Figure 4.

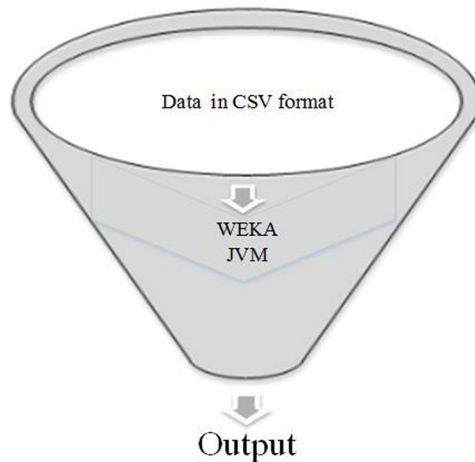


Figure 4: Framework diagram of a system.

4.2 Performance metrics

In continuation, inspired by the productivity of our proposed three-level order conspire, we consider the accuracy performance evaluation measure that mostly utilizes in classification algorithms. Therefore, we consider accuracy in performance metrics:

Accuracy: It is expressed as the ratio of a number of components that are truly classified (TP and TN) with respect to the total number of components (TP, TN, FP, and FN). The mathematical expression for the accuracy parameter is shown in equation (1):

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}. \quad (1)$$

4.3 Performance evaluation

Results are obtained by a series of experimental tests. These tests are also used to analyze the reliability of the proposed three-level DSS. The neural artificial networks (learning algorithm) are the reflection of the BP algorithm. The neural network is constructed and educated by different algorithms. The SMO algorithm (simplest training regimes) was also preferred by the researcher as a support vector machine. The Bayesian networks were also represented by the NB algorithm. This C4.5 algorithm was the protagonist in the analysis. For the standard rule-learning strategy, the RIPPER (JRip) algorithm (most widely used methods of producing rules for classification) was chosen. The 3-NN algorithm (instance-based learner) was configured as a distance metric. Additionally, in the experimental results, the RIPPER, 3-NN, BP, and SMO were referred for voting stands (known as quick voting scheme). The WEKA Computing Toolkit contains all these algorithms. The stratified ten fold cross-validation was used to validate the accurate classification. The split of a dataset into folds is the best example and each fold has the same grade distribution. It presents the whole data. The accuracy of each neural network-based algorithm is illustrated in Table 4 and also shows the relation of three-level classification algorithm performance on a self-recorded database. Each classifier's efficiency is improved by the three-level proposed scheme.

Finally, after the experimental setup, the detailed discussion and conclusion are made in the separate sections.

Table 4: Results of individuals over a dataset

Sr. No.	Classifier	DATA _A		DATA _{AB}	
		Individual (%)	Three-level (%)	Individual (%)	Three-level (%)
1	BP algorithm	80.10	88.10	75.20	88.10
2	SMO algorithm	83.10	89.90	81.70	88.90
3	NB algorithm	74.80	80.20	70.70	78.00
4	C4.5 algorithm	83.20	88.80	83.10	89.80
5	RIPPER (JRip) algorithm	83.70	81.00	85.10	88.60
6	3-NN algorithm	84.20	87.10	77.70	84.30
7	Voting	85.80	88.20	84.40	87.00

Note: The C4.5 uses proper utilization of the proposed model can show up to 88.8% (from Table) accurate results to identify the real reason and gray area of the Pakistani education sector.

5 Discussion

Implementation of intelligent approaches is essential to retrieve data structure and analytical knowledge to explore secret knowledge from the databases of the students. In the global educational age, the application of smart to build accurate and efficient DSSs for evaluating the output of learners is becoming very common as shown in Figure 1. Outstanding feedback has been received on how machine learning tools (MLT) strives to explore different perspectives into learning through new tools and techniques to influence professional activity at all educational levels, as well as corporate learning. The systematic methods of the MTL and how to implement the techniques are already described in Section 3.2.

The usefulness of the academic DSS in evaluating the course curriculum design to improve the quality of the educational curriculum is illustrated. This approach also helps to decrease the failure rate of the students in the academic year. In this research work, the basic principles used in the development and

design of a new set of DSS and also discussed the various methods of evaluating the student performance data for academic decision making.

Zhou et al. [41] studied the precision of six ML algorithms to drop out of a Hellenic Open University distance learning course. They implemented to forecast the academic success of students based on key demographic characteristics, attendance, and their marks in written assignments.

Joshi et al. [19] suggested a guidance framework for the educational decision making system. The implementation of the DSS focuses mainly on enhancing the learning process by designing a detailed curriculum that predicts the characteristics and results of the students. The primary aim of this work to reduce the high rate of poor academic performance among such pupils. The proposed programming depends on a classifier of the neural system which displays more steady conduct and shows preferred exactness over different classifiers. Along the whole line, a user-friendly decision support program was introduced to predict the success of participants, along with a case study on the Physics and English final exams.

We suggest a technique and a basic classification algorithm for finding intelligible student dropout predictive models as rapidly as possible and a multilevel classification model was formulated. They also implemented preprocessing techniques such as data acquisition as well as eliminating the misclassified circumstances from the initial classifier to improve the model's classification accuracy.

5.1 Limitations

The dataset used in the analysis is restricted to limited courses along with a sufficient number of students. Our major focus is on the output performance of the students. The motivation of the students also plays a significant role in predicting their future studies. This cold-war problem is neglected in the prediction of student performance. In addition to these limitations, our predicted model has practical applications for students and universities to increase the retention rate of the students.

6 Conclusion

This research work presents the IDSS based on three-level classifier ML prediction. The computational techniques on educational data mining assist to determine the reason behind the poor performance of Pakistani students in academic years. This novel work first identifies the dataset for experimental analysis about the working of the proposed three-level classifier for the DSS. After supposing and experimentally proving the author introduces a prototype of the software to implement this newly proposed model. The software provides the required results as per the experimental analysis proved about the training of neural networks. The Java-based prototype will be enhanced by the addition of custom features for use in the education sector of Pakistan. The objective and expectation will introduce a new paradigm to identify the reasons for the students who are unable to perform well in the academic years and resulted as a fail candidate. The proper utilization of the proposed model can show up to 88.8% (from Table 4) accurate results to identify the real reason and gray area of the Pakistani education sector. The current scheme results also revealed the accuracy of the different algorithms. Additional features can be quickly applied as per the requirement of the customer and some of the screenshots of this user-friendly software tool. The goal and primary objective of this research work will be used as a guide for decision taking in the curriculum improvement. It also provides the service structure to the student for the improvement of their performance in education.

Acknowledgments: This work was partially supported by the National Natural Science Foundation of China (No. 61572035, No. 61902002, No. 61402011, and No. 61272153), the Natural Science Foundation of Educational Government of Anhui Province of China (No. KJ2016A208), Anhui Provincial Big Data

Foundation (No. 2017032), the Foundation for top-notch academic disciplines of Anhui Province (No. gxbjZD11), the Academic and Technology Leader Foundation of Anhui Province (No. 2019H239), and the Open Project Program of Key Laboratory of Embedded System and Service Computing of Ministry of Education (No. ESSCKF2018-04). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentations.

Conflict of interest: Authors state no conflict of interest.

References

- [1] Abu-Naser SS. A comparative study between animated intelligent tutoring systems AITS and video-based intelligent tutoring systems VITS. 2001. http://dstore.alazhar.edu.ps/xmlui/bitstream/handle/123456789/399/27-05-2019_20.pdf?sequence=1&isAllowed=y.
- [2] Ala'M AZ, Heidari AA, Habib M, Faris H, Aljarah I, Hassonah MA. Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems. In *Evolutionary machine learning techniques*. Singapore: Springer; 2020. p. 11–34.
- [3] Alexeyev A, Solianyk T. Decision-making support system for experts of penal law. In *Data-centric business and applications*. Switzerland: Springer; 2020. p. 163–82.
- [4] Ali S, Ali A, Amin A. The impact of population growth on economic development in Pakistan. *Middle-East J Sci Res*. 2013;18(4):483–91.
- [5] Alloghani M, Al-Jumeily D, Hussain A, Mustafina J, Baker T, Aljaaf AJ. Implementation of machine learning and data mining to improve cybersecurity and limit vulnerabilities to cyber attacks. In *Nature-inspired computation in data mining and machine learning*. Springer; 2020. p. 47–76.
- [6] Baker RSJD, Yacef K. The state of educational data mining in 2009: A review and future visions. *J Edu Data Mining*. 2009;1(1):3–17.
- [7] Bresfelean VP, Ghisoiu N. Higher education decision making and decision support systems. 2009. https://mpr.ub.uni-muenchen.de/26698/1/MPRA_paper_26698.pdf.
- [8] Brusilovsky P. Adaptive educational hypermedia. In *International PEG Conference*. Vol. 10; 2001. p. 8–12.
- [9] Dasgupta J. Imparting hands-on industry 4.0 education at low cost using open source tools and python eco-system. In *New paradigm of industry 4.0*. Springer; 2020. p. 37–47.
- [10] Deniz DZ, Ersan I. An academic decision support system based on academic performance evaluation for student and program assessment. *Int J Eng Edu*. 2002;18(2):236–44.
- [11] Díaz JJS. Artificial intelligence in cardiovascular medicine: Applications in the diagnosis of infarction and prognosis of heart failure. In *Artificial intelligence in precision health*. Spain: Elsevier; 2020. p. 313–28.
- [12] Dutt A, Ismail M, Herawan T. A systematic review on educational data mining. *IEEE Access*. 2017;5:15991–6005.
- [13] Fiebrink R. Machine learning education for artists, musicians, and other creative practitioners. *ACM Trans Comput Edu (TOCE)*. 2019;19(4):1–32.
- [14] García-Saiz D, Zorrilla M. A promising classification method for predicting distance students' performance. In *Educational data mining 2012*. Chania, Greece: Educationa Data Mining; 2012.
- [15] Grivokostopoulou F, Perikos I, Hatzilygeroudis I. Utilizing semantic web technologies and data mining techniques to analyze students learning and predict final performance. In *2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, Rome: IEEE; 2014. p. 488–94.
- [16] Harvey JL, Kumar S. A practical model for educators to predict student performance in K-12 education using machine learning. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, Xiamen, China: IEEE; 2019. p. 3004–11.
- [17] Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU. The rise of “big data” on cloud computing: Review and open research issues. *Inform Syst*. 2015;47:98–115.
- [18] Hormigo IG, Rodríguez ME, Baró X. Design and implementation of dashboards to support teachers decision-making process in e-assessment systems. In *Engineering data-driven adaptive trust-based e-assessment systems*. Greece: Springer; 2017. p. 109–32.
- [19] Joshi KD, Sarker S, Sarker S. Knowledge transfer within information systems development teams: Examining the role of knowledge source attributes. *Decis Support Syst*. 2007;43(2):322–35.
- [20] Kabra RR, Bichkar RS. Performance prediction of engineering students using decision trees. *Int J Comput Appl*. 2011;36(11):8–12.
- [21] Kostopoulos G, Lipitakis AD, Kotsiantis S, Gravvanis G. Predicting student performance in distance higher education using active learning. In *International Conference on Engineering Applications of Neural Networks*. Greece: Springer; 2017. p. 75–86.

- [22] Livieris I, Mikropoulos T, Pintelas P. A decision support system for predicting students' performance. *Themes in Science and Technology Education*. 2016;9(1):43–57.
- [23] Livieris IE. A new ensemble self-labeled semi-supervised algorithm. *Informatica*. 2019;43:2
- [24] Livieris IE, Drakopoulou K, Kotsilieris T, Tampakas V, Pintelas P. Dss-ppsp-a decision support software for evaluating students' performance. In *International Conference on Engineering Applications of Neural Networks*. Greece: Springer; 2017. p. 63–74.
- [25] Livieris IE, Drakopoulou K, Mikropoulos TA, Tampakas V, Pintelas P. An ensemble-based semi-supervised approach for predicting students' performance. In *Research on e-learning and ICT in education*. Switzerland: Springer; 2018. p. 25–42.
- [26] Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: An applied review. *Int J Remote Sens*. 2018;39(9):2784–817.
- [27] Mosavi A, Rabczuk T, Varkonyi-Koczy AR. Reviewing the novel machine learning tools for materials design. In *International Conference on Global Research and Education*. Romania: Springer; 2017. p. 50–8.
- [28] Mothilal RK, Sharma A, Tan C. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Barcelona, Spain: Association for Computing Machinery (ACM); 2020. p. 607–17.
- [29] Naidu VR, Singh B, Al Farei K, Al Suqri N. Machine learning for flipped teaching in higher education—a reflection. In *Sustainable development and social responsibility*, vol. 2, Switzerland: Springer; 2020. p. 129–32.
- [30] Niet YV, Díaz VG, Montenegro CE. Academic decision making model for higher education institutions using learning analytics. In *2016 4th International Symposium on Computational and Business Intelligence (ISCBI)*. Olten, Switzerland: IEEE; 2016. p. 27–32.
- [31] Noaman AY, Luna JM, Ragab AHM, Ventura S. Recommending degree studies according to students' attitudes in high school by means of subgroup discovery. *Int J Comput Intell Syst*. 2016;9(6):1101–17.
- [32] Osmanbegovic E, Suljic M. Data mining approach for predicting student performance. *Econ Rev: J Econ Business*. 2012;10(1):3–12.
- [33] Rashad HM, Aly WM, Hegazy OF. An educational data mining system for advising higher education students IJIE. 2013;7(10).
- [34] Romeo L, Loncarski J, Paolanti M, Bocchini G, Mancini A, Frontoni E. Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. *Expert Syst Appl*. 2020;140:112869.
- [35] Romero C, Ventura S. Educational data mining: A survey from 1995 to 2005. *Expert Syst Appl*. 2007;33(1):135–46.
- [36] Romero C, Ventura S. Educational data mining: A review of the state of the art. *IEEE Trans Syst, Man, and Cybernet, Part C (Appl Rev)*. 2010;40(6):601–18.
- [37] Scarpazza C, Baecker L, Vieira S, Mechelli A. Applications of machine learning to brain disorders. In *Machine learning*. Netherland: Elsevier; 2020. p. 45–65.
- [38] Soliman H, Tabak F. Deep learning framework for RDF and knowledge graphs using fuzzy maps to support medical decision. *J Int Res Med Pharma Sci*. 2020;14(3):92–7.
- [39] Vo TNC, Nguyen HP. A knowledge-driven educational decision support system. In *2012 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future*. United States: IEEE; 2012. p. 1–6.
- [40] Yadav SK, Bharadwaj B, Pal S. Data mining applications: A comparative study for predicting student's performance, arXiv:<http://arXiv.org/abs/arXiv:1202.4815>, 2012.
- [41] Zhou Y, Zheng S, Zhang G. Machine learning-based optimal design of a phase change material integrated renewable system with on-site pv, radiative cooling and hybrid ventilations—study of modelling and application in five climatic regions. *Energy*. 2020;192:116608.