

Ensemble Techniques in E-Learning Student's Performance

Subjects: Computer Science, Information Systems

Contributor: Abdulkream A. Alsulami, Abdullah S. AL-Malaise AL-Ghamdi, Mahmoud Ragab

Educational institutions have dramatically increased in recent years, producing many graduates and postgraduates each year. One of the critical concerns of decision-makers is student performance. Educational data mining techniques are beneficial to explore uncovered data in data itself, creating a pattern to analyze student performance.

Keywords: educational data mining ; student performance ; classification techniques

1. Introduction

One of the main concerns for educational institutions is to analyze the factors that affect student performance. Every school tries to reduce the failure of their students. The most popular technique to evaluate and predict students' performance is educational data mining (EDM) ^[1]. EDM is about developing methods to deal with the different types of data in educational systems to improve students' learning outcomes ^[2]. EDM creates and modifies statistical, machine learning, and data mining approaches. EDM's primary objective is information extraction from educational data for educational decision-making. Educational data mining (EDM) can predict students' academic achievement early ^[3]. Their use could enhance the analysis of students' learning processes while taking into account how they interact with the environment.

2. Ensemble Techniques in E-Learning Student's Performance

In ^[4], researchers aimed to predict the student's success on the exam. They modeled the study by using the decision tree and K-nearest neighbor. The study concluded that the decision tree predicts students' pass or fail status in an academic course with the best results. In ^[5], researchers compared different classification techniques: NaiveBayesSimple, multilayer perceptron, SMO, J48, and REP-tree—in the field of comparing student performance. The data set was collected from a computer science department of a college with 300 student records. WEKA was the tool to use in the study. From the results, researchers concluded that the performance of multilayer perceptron is the most effective algorithm for predicting student performance. In comparison to other algorithms, multilayer perceptron accuracy was higher than other classifiers. In ^[6], researchers aimed to determine the basic factors that have a significant effect on secondary student performance. To do so, they combined single and ensemble-based classifiers to create the proper classification model, which they then used to forecast academic success. In the beginning, three data mining methods were used: decision tree, multilayer perceptron (MLP), and a PART; moreover, three ensemble techniques (multi-boost, bagging (BAG), and voting) were used individually. To improve the previous classifiers' performance, a single classifier and an ensemble classifier were combined, generating nine new models. According to the evaluation's findings, multi-boost with MLP outperformed other approaches in terms of accuracy. In ^[7], using data mining techniques, researchers were trying to predict student dropout. The findings demonstrated the possibility of dropout with accuracy rates greater than 0.80 in most situations and false positive values varying from 0.10 to 0.15 on average. K-nearest neighbors, random forest, support vector machines, decision trees, logistic regression, and naive Bayes were among the methods they were contrasting. Random forest outperformed other machine learning techniques in terms of accuracy, F-measure, and precision. In ^[8], researchers have concentrated on forecasting student performance in several interactive online sessions by exploring the information gathered using the E-learning and design suite. The data set keeps track of student participation during classes, including text editing, keystrokes, and the amount of time spent on each assignment. They used five well-known classifiers: naive Bayes, random forest, support vector machine, multi-layer perception, and logistic regression. Three distinct evaluation methods were utilized: five-fold cross-validation and random data split for training and testing. The model was trained in all sessions except the one used for testing. According to the results, the RF classifier model obtained the best accuracy. In ^[9], researchers investigated various classifier algorithms that are proposed to predict secondary school students' success in mathematics and Portuguese lessons. They classified using support vector machine (SVM), linear discriminant

analysis (LDA), and K-nearest neighbor (KNN). Their experimental results demonstrated that the SVM method performed better for the unbalanced class distribution problem. In ^[10], the classification technique being evaluated by researchers was a hybrid classification. To do so, they used the radial basis function network, C4.5, random forest, and multilayer perceptron algorithms. They observed that hybrid classification algorithms perform more accurately than single algorithms. In ^[11], researchers were clustering the data by using the K-nearest neighbors (KNN) algorithm with the help of Harris hawks optimization (HHO). Once they classify all of the solutions, redistribution for the solutions into a search space will be applied. Several different machine learning classifiers were used to validate the overall prediction system, such as naive Bayes, KNN, LRNN, and artificial neural network. The results collected demonstrate the significance of anticipating student performance early to reduce student failure and enhance the overall effectiveness of the educational institution. Furthermore, given that LRNN is a deep learning method that can observe past and current input values, the results showed that the modified HHO and LRNN combination outperforms other classifiers with an accuracy of 0.92. In ^[12], Researchers concentrated on how crucial it was to take advantage of both technological advancements and potential educational contributions. They tested a new PFA strategy based on various ensemble learning techniques to improve the forecasting of student performance. (random forest, AdaBoost, and XGBoost). The results have demonstrated that XGBoost could predict future student acquisition with the highest performance. In ^[13], researchers presented the data mining technique used to forecast first-year students' academic performance. They chose three different data models for learning stages and tested them based on the dates of entry, end of the first, and end of the last semesters. Records of bachelor students who enrolled in a program offered by the institution between 2006/2007 and 2015/2016 were obtained and gathered through the institutional database. The best overall performance was gained by a support vector machines (SVM) model, which was chosen to perform database sensitivity analysis. **Table 1** shows some papers that used different data mining techniques in order to predict the performance of students.

Table 1. Comparison of data mining techniques in predicting student's performance.

Year, Author(s)	Methodology	Key Findings
2022, Aremu, Dayo Reuben, Awotunde, and Ogbuji ^[4]	Decision tree (DT) and K-nearest neighbor (KNN)	Decision tree DT for predicting the pass/fail status of students delivers the most successful outcomes
2021, Siddique, Ansar, et al. ^[6]	MLP, J48, and PART BAG, MB, and VT	Concluded MultiBoost with MLP outperformed the others.
2021, Palacios, Carlos A., et al. ^[7]	DT, KNN, LR, NB, RF, and SVM	RF algorithm ranked first among the others.
2022, Begum, Safira, and Sunita S Padmanavar ^[8]	KNN, LDA, and SVM	Shown that the SVM is for the unbalanced class distribution problem.
2022, Brahim, Ghassen Ben ^[9]	MLP, RF, SVM, NB, and LR	Showed the best classification accuracy performance
2021, Kumar, A. Dinesh, R. Pandi Selvam, and V. Palanisamy ^[10]	multilayer perceptron, Radial basis function network, C4.5, and random forest algorithm	Hybrid classification algorithms are more accurate than individual classification algorithms.
2021, Gil, Paulo Diniz, et al. ^[13]	DT, RF, SVM, and ANN	SVM has the highest accuracy among others
2022, Joshi, Manuj, and Chawda ^[14]	NaiveBayesSimple, Multilayer Perception, SMO, J48, and REPTree	Concluded multilayer perception algorithm is most appropriate for predicting student performance.
2021, Ahammad, Khalil, et al. ^[15]	support vector machine, naive Bayes, K-nearest neighbours, XG-boost, and multi-layer perceptron	Multi-layer perceptron achieved the highest accuracy

In ^[16], researchers attempted to determine the factors influencing academic performance. Thus, they made use of two different types of data sets. The first data set demonstrates how the performance in a course's required courses might affect a student's performance in the current course. The second data set suggested that the student's grade in any course is related to their performance in the semester until the midterm test. In ^[17], the results of the model showed that the main contribution to predicting academic performance is related to the following factors: interview, task, questionnaire, and age. The access factor measures student's access to the module, including access to forums and glossaries. Questionnaire factors summarize the variables in the questionnaire related to the visit and the attempt. The age factor contains the student's age. In ^[18], the study aimed to investigate the factors affecting student performance. Researchers reviewed and analyzed 36 articles. They concluded that the performance in previous classes and grades, the students' e-

Learning activities, and their demographic background had an impact on the performance of the student, academically speaking. In order to determine whether students' learning behaviors were important, researchers examined the same data set [19]. They used the ensemble methods, voting, bagging, and boosting, alongside traditional data mining methods, support vector machines, decision tree (ID3), K-nearest neighbor, and naïve Bayes. With the help of the voting process, the highest accuracy was achieved. In [14], an investigation was conducted on learners' relationships with e-learning. A combination of ensemble algorithms with three different types of classifiers was used: decision trees, K-nearest neighbors, and support vector machines. It was found that learners' features were strongly correlated with their performance in the study. In contrast, ensemble techniques increased accuracy. In [15], in order to help decision-makers make the best choices for their organizations, researchers used ensemble methods to predict student performance. They used naïve Bayes, decision tree, and K-nearest neighbor methods. The voting technique was used to combine the three methods. In most scenarios, the proposed model improved the accuracy of naïve Bayes.

References

1. De Andrade, T.L.; Rigo, S.J.; Barbosa, J.L.V. Active Methodology, Educational Data Mining and Learning Analytics: A Systematic Mapping Study. *Inform. Educ.* 2021, 20, 2.
2. Cristobal, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 2020, 10, e1355.
3. Liñán, L.C.; Pérez, Á.A.J. Educational Data Mining and Learning Analytics: Differences, similarities, and time evolution. *Int. J. Educ. Technol. High. Educ.* 2015, 12, 98–112.
4. Aremu, R.D.; Awotunde, J.B.; Ogbuji, E. Predicting Students Performance in Examination Using Supervised data mining techniques. In *Proceedings of the Informatics and Intelligent Applications: First International Conference, ICIIA 2021, Ota, Nigeria, 25–27 November 2021*.
5. Hassan, Z.; Braendle, U.; Farah, A. Enhancing prediction of student success: Automated machine learning approach. *Comput. Electr. Eng.* 2021, 89, 106903.
6. Siddique, A.; Jan, A.; Majeed, F.; Qahmash, A.I.; Quadri, N.N.; Wahab, M.O.A. Predicting Academic Performance Using an Efficient Model Based on Fusion of Classifiers. *Appl. Sci.* 2021, 11, 11845.
7. Palacios, C.A.; Reyes-Suárez, J.A.; Bearzotti, L.A.; Leiva, V.; Marchant, C. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy* 2021, 23, 485.
8. Begum, S.; Padmanavar, S.S. Genetically Optimized Ensemble Classifiers for Multiclass Student Performance Prediction. *Int. J. Eng. Trends Technol.* 2022, 15, 223–235.
9. Brahim, G. Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features. *Arab. J. Sci. Eng.* 2022, 47, 10225–10243.
10. Kumar, A.D.; Selvam, R.P.; Palanisamy, V. Hybrid classification algorithms for predicting student performance. In *Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Online, 25–27 March 2021*.
11. Turabieh, H.; Azwari, S.A.; Rokaya, M.; Alosaimi, W.; Alharbi, A.; Alhakami, W.; Alnfai, M. Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. *Computing* 2021, 103, 1417–1438.
12. Amal, A.; Khaldi, M.; Aammou, S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. In *Interactive Learning Environments*; Taylor & Francis: Abingdon, UK, 2021; pp. 1–20.
13. Gil, P.D.; da Cruz Martins, S.; Moro, S.; Costa, J.M. A data-driven approach to predict first year students' academic success in higher education institutions. *Educ. Inf. Technol.* 2021, 26, 2165–2190.
14. Joshi, M.; Chawda, N.S. Implementation of Data Mining Techniques in Predicting Selection Chances in Competition. In *Proceedings of the Sixth International Congress on Information and Communication Technology: ICICT*; Springer: Singapore, 2022; Volume 4.
15. Ahammad, K.; Chakraborty, P.; Akter, E.; Fomey, U.H.; Rahman, S. A comparative study of different machine learning techniques to predict the result of an individual student using previous performances. *Int. J. Comput. Sci. Inf. Secur. (IJCSIS)* 2021, 19, 1.
16. Nahar, K.; Shova, B.I.; Ria, T.; Rashid, H.B.; Islam, A.S. Mining educational data to predict students' performance. *Educ. Inf. Technol.* 2021, 26, 6051–6067.
17. Pu, H.-T.; Fan, M.-Q.; Zhang, H.-B.; You, B.-Z.; Lin, J.-J.; Liu, C.-F.; Zhao, Y.-Z.; Rui, S. Predicting academic performance of students in Chinese-foreign cooperation in running schools with graph convolutional network. *Neural*

Comput. Appl. 2021, 33, 637–645.

18. Abu Saa, A.; Al-Emran, M.; Shaalan, K. Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive data mining techniques. *Tech. Knowl. Learn.* 2020, 24, 567–598.
19. Mrinal, P.; Taruna, S. An ensemble-based decision support system for the students' academic performance prediction. In *ICT Based Innovations: Proceedings of CSI 2015*; Springer: Singapore, 2018.

Retrieved from <https://encyclopedia.pub/entry/history/show/124568>