# Machine learning for learning personalization to enhance student academic performance

Juan Carlos Muñoz-Carpio[1,2], Zohaib Jan[3] and Angelo Saavedra[4]

[1] *Central Queensland University, 160 Ann Street, Brisbane, Australia*
[2] *Torrens University, Fortitude Valley, Brisbane, Australia*
[3] *University of South Australia, Mawson Lakes, Adelaide, Australia*
[4] *University of New England, Elm Avenue, Armidale, Australia*

**Abstract**

Machine learning, and artificial intelligence have allowed assessment of large and complex datasets for various analytical purposes such as predicting and forecasting, segmenting, object detection *etc*. They influence many sectors and industries, including the field of education in identifying whether a student's engagement and learning performance can impact their academic success. These improvements are creating new teaching and learning strategies to enhance students' performance and their overall education. Since datasets are prone to randomness and noise and are generally unbalanced, which is especially true for academic datasets. Therefore, this hinders the learning capabilities of a machine learning model. In this paper, we propose a new performance prediction model using an optimized ensemble classifier which is a type of a machine learning model for predicting students' learning performance using and unbalanced datasets. The results are compared with existing state-of-the-art ensemble methods, including bagging, and boosting, currently used in the literature. The results of the proposed model using a student dataset reveals an accuracy of more than 80%.

**Keywords**

Machine learning, student performance prediction, learning performance, educational data, ensemble classification

## 1. Introduction

Demand for the use of machine learning algorithms in predictive analyses has increased since the explosion of data in real-world scenarios. In particular, the shift to online learning due to the recent pandemic was inevitable and therefore, new learning paradigms were developed. As such, identifying key factors or predicting students' academic performance given a set of background information will prove beneficial not only for the instructors but also for the learners. The current learning management system is used to support the processes in formal learning settings and so needs to be updated to match present-day requirements. Well known examples include the open-source packages Moodle and Sakai,

*XVI LATIN AMERICAN CONFERENCE ON LEARNING TECHNOLOGIES, October 19–21, 2021, Arequipa, Perú*
EMAIL: j.munoz@cqu.edu.au (A. 1); muhammad.jan@unisa.edu.au (A. 2); msaavedr@une.cedu.au (A. 3)
ORCID: 0000-0003-0251-5510 (A. 1);0000-0002-5066-4118 (A. 2); 0000-0002-2750-3113 (A. 3)

as well as commercial products such as the Blackboard system used by universities worldwide. Machine learning and its applications, particularly in learning personalization, are gaining popularity, including learning assessments to improve students' outcomes through design assessment and feedback [1]. Some of this specialist focus on information classification, student competencies and retrieval technique in natural language processing, well known as report-based method and metanalysis, is used to identify a person's competences [2,3]. The need for machine teaching forces the knowledge to be externalized into content, also frequently called learning material, in the form of web-based training material. Knowledge in traditional enhanced learning approaches and learning material content can be considered standardized, externalized, and structured. However, data of the learning process from problem solvers and their learning actions can be measured, assembled, scrutinized, and reported using learning analytics (LA). In this regard LA contributes to enhance both the learning process and the environment in which it occurs. It provides appropriate and timely feedback regarding learning processes to stakeholders (teachers, administrators, parents, and students) [4]. For example, an approach towards structured knowledge can also include the use of videos and haptic tools that can be employed as part of the process-oriented learning. The use of personalized learning materials as resources that may be provided to learners in the most efficient way to help them learn more effectively [5].

In this instance, content, learning challenges and activities can be created as sources to facilitate learning processes and therefore to augment understanding and enhance learning. Elaboration of cognitive learning models can be applied to predict students' learning behaviors toward the use of technology-enhanced learning. According to [6,7] the cognitive architecture that has been applied to predict student's behavior as they use technology-enhanced learning to model the cognition of the learner from their interaction with the system. The design of a collaborative learning approach must enable learning immersion in a scaffolding manner; this cannot be achieved by merely providing a set of learning tools or collaborative group tasks to enhance learners' competency [8]. According to [9], there are five conditions for making collaborative learning superior than competitive learning or individual learning: (1) positive interdependence (everyone is working towards the same goals), (2) individual accountability/personal responsibility (everyone is responsible for themselves), (3) interaction promotion (interaction that is mostly face-to-face), (4) interpersonal and small group skills (use of communication abilities to collaborate effectively and perform well as part of a team), and (5) to increase the efficacy of the group, it should be evaluated often and on a regular basis. In this instance, the collaborative learning design uses a method that examines collaboration and student experience with no tutor intervention [10]. In this context, machine learning can be used as it can support an automated process of analysis to monitor learning performance. Figure 1 is a visualization of the relationship between traditional technology-enhanced learning approaches and machine teaching.

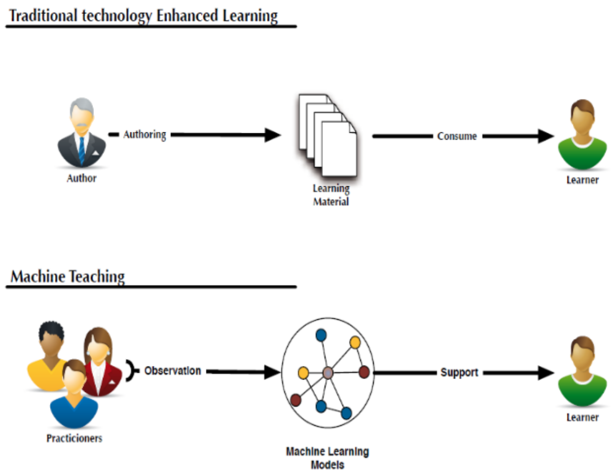


**Figure 1.** Machine teaching used as an alternative compared to technology enhanced teaching [11]

In the field of predictive analytics, ensemble classifiers [12] are a popular machine learning methodology that aim to improve single classifiers classification performance by fusing together multiple classifiers. By taking advantage of perturb and combine [13], single classifier models are outperformed by ensemble classifiers. Ensemble classifiers use an approach known as the random subspace method [14,15] to perturb a given input, in which random sub-samples of the input data are created to train a large number of classifiers on. The students' academic dataset, as originally proposed by [16], is generally divided into three classes. The sample distributions of students belonging to different grade classes is not the same; as a result, any prediction model developed using such data will be skewed. In this study we present a new ensemble classifier methodology that will mitigate the class imbalance through the incorporation of clustering and optimization.

The purpose of this research is to investigate the use of machine learning approaches for predictive analytics such ensemble classifiers to predict a student's academic performance based on demographic, interaction, class participation and various other features. The contributions of this paper are as follows:

- A methodology of utilizing ensemble classifiers for predicting academic performance
- A novel methodology of optimizing machine learning model on biased data
- Experimental analysis on academic as well as benchmark datasets

## 2. Background

This section discusses the application of machine learning in education, specifically the different machine learning techniques used to predict student performance and success. This is particularly important because predicting each student's performance will enable educators to identify potential poor performers and providing further assistance to the students at risk [17]. Such support can take the form of additional learning activities, resources and learning tasks [18]. In fact, identifying students who are more likely to drop out of classes early allows for the timely deployment of support mechanisms to keep these students from dropping out. [19]. Moreover, this information is useful for the successful implementation of student retention strategies, which directly affects graduation rates

[19]. To predict students' performance, the authors used decision trees, neural networks, naïve Bayes method, instance-based learning algorithms, logistic regression, and support vector machines.

Decision tree is a non-parametric supervised learning method used for classification and regression. The objective is to learn basic decision rules derived from data attributes to forecast the value of a target variable. A decision tree algorithm seeks for the most efficient way to divide data into portions that are as homogenous as feasible [20]. For example, [21] built a model that uses decision tree algorithms to help students in an introductory programming course predict their anticipated final scores. Decision trees have several advantages, including being simple to process, requiring minimal data preparation, handling both numerical and categorical data, and performing well even when the underlying model from which the data were created violates some assumptions. In addition, the model can be validated using statistical testing, in this way accounting for the reliability of the model.

Another type of inductive learning is artificial neural networks. They are based on computer models of biological neurons and neural networks that are similar to the human central nervous system [17]. These networks are densely interconnected and have an intrinsic proclivity for learning from experience as well as uncovering new information. [18]. Classification occurs in two independent stages. First, to identify the input-output mapping, the network is first trained on a collection of paired data. The network is then utilized to determine the classifications of a fresh batch of data once the weights of the connections between neurons have been fixed. [17,18]. Due to the self-learning and self-adapting features, this method has been effectively used to address complex real-world problems [18]. For example [18], built a user-friendly software solution for forecasting the performance of students enrolled in a secondary school mathematics class (Lyceum) in Greece using neural network classifiers. They determined that their model was more consistent and produced better classification results than the other classifiers (e.g. decision trees, Bayesian networks, classification rules and support vector machines).

The "naïve" assumption of conditional independence between any pair of features given the value of the class variable underpins a set of supervised learning techniques based on Bayes' theorem. This algorithm captures the assumption that every attribute is independent from other attributes given the state of the class attribute [17]. Although it is considered the simplest form of a Bayesian network [22], In many real-world contexts, such as document categorization and spam filtering, naïve Bayes classifiers have performed well. In educational settings, naïve Bayes was used in combination with classifications [4] and a decision tree model [20] to predict the students' academic success. Instance-based learning algorithms derive from the nearest neighbor pattern classifier [23]. They are highly similar to the modified nearest neighbor algorithms, which store and use only a few occurrences to make classification predictions [23]. In contrast to the non-incremental edited closest neighbor method, which has the primary purpose of keeping consistency with the initial training set, instance-based learning algorithms are incremental, and their goal is to maximize classification accuracy on future given cases [23].

Logistic regression is a linear model commonly used to predict student success. For example, it has been used by the Noel-Levitz Corporation in the United States to identify new students' chances of withdrawal based on their records and known entry characteristics (e.g. sex, age and previous qualifications) [24]. It is predicted that some elements would have a considerably bigger effect on students' chances of success than others, the study generates an algorithm weighted for various aspects. Likewise, at Napier University in the UK, a logistic regression model indicated a link between dropping out and working while a student – students who worked more than 15 hours per week had a greater likelihood of dropping out. As a result, it may be advised to these students that they limit their working hours [24].

Support vector machines are a type of supervised learning algorithms that may be used for classification, regression, and outlier identification. [25] proposed an ensemble support vector machine model based on almost 100 features, including psychological educational factors, for predicting a student's graduation. Using data from a state university in the US, the model turned out to be effective in predicting students' graduation with a high level of accuracy and precision. Another example is [26], who used and compared four data mining methods to predict student failure: decision tree, random forest, neural network, and support vector machine. In this paper, we propose a clustering-based ensemble as a contribution to the research. By applying a clustering ensemble, we expect a more accurate prediction of students' academic performance. In terms of consistency, dependability, and accuracy, a successful clustering ensemble should be able to outperform the individual clustering methods [27,28].

## 3. Proposed Methodology

The proposed ensemble learning framework starts by generating data sub-samples through cluster centroid methodology and then trains a collection of different base classifiers on all created sub-samples. This results in generating a pool of trained base classifiers that is represented as a binary combinatorial problem that is optimized to choose the optimum subset of classifiers that can maximize ensemble accuracy.

### 3.1. Sub-samples generation

Due to the presence of randomness and noise, the datasets are not perfectly balanced, meaning that the number of samples are not evenly distributed across the different classes. Consequently, this will affect the training process of not only single classifiers but also the ensemble of classifiers. Any classifier that is trained on an unbalanced dataset will be biased towards the majority class, therefore affecting the generalization performance of the classifier. Ensemble classifiers generally train a multitude of classifiers to generate the base classifier by employing subsampling techniques. For example, bagging is a common strategy used to generate bags of input data and train multiple classifiers on generated bags. However, if the dataset is biased or unbalanced, the generated bags will also be biased. Therefore, to avoid this issue in this study, we utilize a cluster centroid method of under-sampling. The centroids of generated data clusters of the majority class are used to match the number of samples of the minority class. This causes the majority class, which is essentially overwhelming the minority class, to be under-sampled without losing a large portion of critical information.

### 3.2. Base classifier pool generation

A group of different base classifiers (e.g., artificial neural networks, support vector machine, decision tree, K-nearest neighbor, and naïve Bayes) is trained on the under-sampled, optimally balanced dataset to form the base classifier pool. These classifiers are distinct in their nature and carry with them a variety of learning capacities.

### 3.3. Base classifier pool optimization

Selecting the optimal subset of base classifiers from a generated pool of trained base classifiers is referred to as a binary combinatorial problem. It is established in research that binary combinatorial problems are NP-hard problems, especially for a large search space. Therefore, in this study the base classifier pool is represented as a binary combinatorial optimization problem. For this purpose, binary particle swarm optimization (BPSO) is utilized as a black box tool to optimize the pool of trained base classifiers. BPSO takes in a set of candidate solutions (a subset of base classifiers) and tries to find the best solution that can maximize the generalization ability of the ensemble using an updated position/velocity update method. The problem for optimization is formulated as follows:

$$minimize\ (f(\xi))\ subject\ to\ \xi\ \in bcp \tag{1}$$

where $\xi$ is a possible ensemble solution consisting of a subset of a base classifier from the pool of trained base classifiers $bcp$ and $f(\xi)$ is the objective function/cost function of the optimisation process, given as:

$$f(\xi)\ =\ (TP+TN)/(TP+FP+TN+FN) \tag{2}$$

where TP is the true positive score, TN is the true negative score, FP is the false positive score, and FN is the false negative score. For a dataset $X=\{(x_1,y_1),\{(x_2,y_2),\dots\{(x_n,y_n)\}$ containing d-dimensional feature vectors $x\in\mathbb{R}^d$, each associated with a discrete class label $y\in\{1,2,3,\dots,V\}$. These scores are calculated using the predicted class labels $y'$ of the ensemble $\xi$ generated by using the validation data set as input for each of the classifiers $\zeta$ in the subset:

$$\xi=\{\zeta(x)^1,\zeta(x)^2,\dots,\zeta(x)^{bcp}\} \tag{3}$$

The mode of the predictions is taken to generate the final ensemble solution, which depicts majority voting and is given as:

$$y'=mode(\xi) \tag{4}$$

## 4. Experiments and Analysis

To analyze the efficacy of the proposed optimized ensemble classifier, a dataset that categorizes student academic performance was used [16] given demographic features, academic background, parents' participation, and behavior attributes is used (please refer to the relevant publication for further information). The dataset has three prediction classes, H (higher distinction), M (medium distinction), and L (lower distinction). There are 142 samples from H group, 211 from M group and 127 from L group, so there is a clear imbalance in the sample distribution of various classes and the dataset is biased towards M group. A snapshot of the student academic dataset is given in Table 1.

A 10-fold cross validation was performed to incorporate randomization, and classification accuracy over 10 folds was averaged and reported. The proposed ensemble learning framework was implemented in Python using default implementation of the base classifiers from the ScikitLearn library and cluster centroid method from the Imbalance learning library.

**Table 1**
Description of student dataset used in the experimentation.

| Attribute / predictor variables | Attribute value |
|---|---|
| Gender | M / F |
| Nationality | Jordan, Kuwait, Lebanon, Saudi Arabia, Iran, USA, Egypt |
| Place of birth | Jordan, Kuwait, Lebanon, Saudi Arabia, Iran, USA, Egypt |
| Stage id | Lower Level, Middle School, High School |
| Grade id | G-02 - G-12 |
| Section id | A / B/ C |
| Topic | Math / English / IT / Arabic / Science / Quran |
| Semester | First year / Second year |
| Relation | Primary caretaker of the student: Father / Mother |
| Raised hands | # of times student interacted in the class: 0 - 100 |
| Visited resources | # of times student visited the resources provided: 0 - 99 |
| Announcements view | # of times student viewed the announcements made: 0 - 99 |
| Discussion | # of times student discussed potential issues among peers on forums etc.: 0 - 99 |
| Parent answering survey | Yes / No |
| Parent school satisfaction | Good / Bad |
| Student absence days | Above 7 / Under 7 |
| Grade Class | Low / Medium / High |

Some of the variables in Table 1 such as the number of times student interacted in the class or visited the e-resources are collected using the learning management system called Kalboard 360 e-learning system.

## 4.1. Results and analysis of student’s dataset

As stated earlier the number of samples is not the same from different classes causing biasness in the predictive performance of learning classifiers. Therefore, to mitigate the issue the cluster centroid method discards almost 10% of the samples from the majority class to accommodate for the minority classes. This is done so that the majority class does not overwhelm the minority data class causing a biasness towards the majority class. Figure 2 below shows the sample distribution of various classes in 3 different clusters after conducting a cluster centroid under sampling. It can be noted that after under-sampling the majority class approximately equal number of samples are present in each data cluster.

**Table 2**
Comparative analysis of the proposed ensemble learning framework and legacy ensemble approaches on student's dataset

| Dataset | Proposed approach | Bagging [E.A. Ameri 2016] | Boosting [E.A. Ameri 2016] | Random Forest [E.A. Ameri 2016] |
|---|---|---|---|---|
| *Students* | 81.28 | 75.6 | 79.1 | 75.6 |

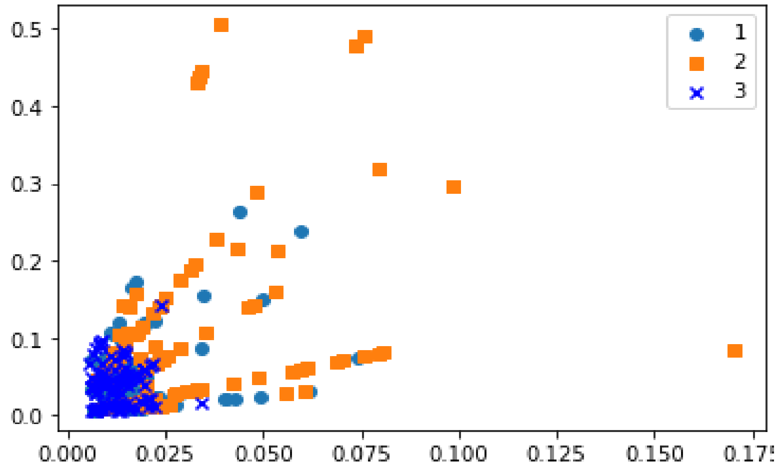


**Figure 2.** Cluster centroid analysis of the 3 data classes in the student's academic performance dataset

The average classification accuracy over 10- folds of the proposed approach is given in Table 2. It can be noted from Table 2 that the proposed ensemble approach achieved higher classification accuracy than the legacy ensemble classifiers which were initially used to classify the dataset.

As shown in Table 2, the proposed ensemble approach achieved higher classification accuracy than legacy ensemble classifiers and is more appropriate for imbalance datasets. This is a clear indication that as we progress deeper into the information age the amount of data is and will increase exponentially. Consequently, the field of predictive analytics will be relied upon more and more. However, the curse of dimensionality, noise and randomness will continue to plague the data that is generated, and existing models need to be revised accordingly to leverage the power of machine learning and availability of data to assist in facilitating a more conducive learning environment. Therefore, the existing learning management systems can leverage on the power of machine learning models to identify students in the system and flag students that will need further assistance or help before they show poor academic performance.

## 4.2. Results and analysis on UCI dataset benchmark dataset

Nine machine learning benchmarking classification datasets from the University of California Irvine repository were utilized to further examine the performance of the suggested ensemble technique in this study. The specifics of these datasets are shown in Table 3 below.

**Table 3**
UCI benchmarking classification datasets used in experimentation

| Dataset | # of samples | # of features | # of classes |
|---|---|---|---|
| *Diabetic* | 768 | 8 | 2 |
| *Ecoli* | 336 | 7 | 8 |
| *Ionosphere* | 351 | 33 | 2 |
| *Iris* | 150 | 4 | 3 |
| *Liver* | 345 | 6 | 2 |
| *Segment* | 2310 | 19 | 7 |
| *Sonar* | 208 | 60 | 2 |
| *Vehicle* | 946 | 18 | 4 |
| *Wine* | 178 | 13 | 3 |

## 4.3. Comparative analysis

The average classification accuracy is collated and compared with existing state-of-the-art ensemble classifier techniques [29]. The classification accuracies are derived from the relevant studies and are given in Table 4, with the greatest classification accuracies in bold.

**Table 4**
Comparative analysis of the proposed ensemble classifier with OEC-ILC, bagging and boosting

| Dataset | Proposed approach | OEC-ILC | Bagging | Boosting |
|---|---|---|---|---|
| *Diabetic* | 0.7588 | 0.7734 | 0.7602 | 0.7462 |
| *Ecoli* | 0.9050 | 0.8564 | 0.8867 | 0.8890 |
| *Ionosphere* | 0.9575 | 0.9157 | 0.9136 | 0.9000 |
| *Iris* | 0.9750 | 0.9600 | 0.9667 | 0.9733 |
| *Liver* | 0.7786 | 0.7127 | 0.7024 | 0.7071 |
| *Segment* | 0.9443 | 0.9950 | 0.9680 | 0.9572 |
| *Sonar* | 0.9059 | 0.9080 | 0.8551 | 0.8266 |
| *Vehicle* | 0.8482 | 0.8100 | 0.8424 | 0.8096 |
| *Wine* | 0.9595 | 0.9813 | 0.9778 | 0.9722 |

It can be shown that the suggested ensemble classifier generated performance increases of 1.33% over OEC-ILC, 1.77% over bagging, and 2.79% over boosting. Thus, adding to the fact that the proposed approach can not only be effective for academic datasets but other unbalanced datasets as well.

## 5. Discussion

This study proposed an ML-based model for predicting student's academic performance. The same model was tested on a real-world dataset as well as benchmark datasets. Due to noise and randomness the datasets are biased and most of the times having more samples from the class that a user is not interested in. Ensemble classifier models are known to be effective when there is a bias in a dataset because they control the bias and variance by employing various strategies. Therefore, this study proposed and tested the efficacy of ensemble-based models using a real-world dataset.

The proposed model can be embedded in existing Learning Management System (LMS), that will assist the teaching staff to focus more on "flagged" students. This will allow the LMS to proactively "infer" using the data features mentioned before to predict a student's grade before they have participated. We expect that the results obtained from this analysis will assist to identify learning needs and learners' performance. Learners can be supported with a variety of multiple learning material representations targeted to specific learning needs. This approach is essential when students are learning new problem domains, abstract concepts or new theories that may include dynamic processes for learning. This conjunction of machine learning and student's demographic and class participation data contributes to LA. Since majority of education institutes are relying more and more on digital education, thus, creating a multitude of data that is not usually analyzed or processed for various reasons. LA can assist in not only identify student's performance but also assist in evaluating a course's performance to better understand the learning implications in a more elaborated manner.

## 6. Conclusion

Academic learning performance is a major problem for many academic institutions and universities, and if not addressed appropriately, it may cause substantial distress, poor academic performance, and increased dropout rates. Particularly, in terms of providing educational frameworks aligned with delivering learning resources and improving student's academic performance looking at the problem from multiple angles and in a multidimensional manner. By using an ensemble classifier, a computable training model using students' dataset was identified. In this manner, learning competencies are first determined and subsequently optimized. The data used contain hidden information that could be used to determine the steps for a student's academic achievement. In this paper, a new performance prediction model for a binary combinatorial optimization problem based on learned base classifiers is provided. Further research will be conducted in the future to employ more assessment methodologies to explore the links between different features. Also, to determine which characteristics are more important than others in influencing a student's overall academic achievement. Further research may be conducted to investigate patterns in other educational systems, which will aid in the improvement of the LMS.

## References


[1] Sekeroglu, B., Dimililer, K., and Tuncal, K. 2019. "Student Performance prediction and Classification Using Machine Learning Algorithms," in Proceedings of the 2019 8th International Conference on Educational and Information Technology. Cambridge, UK. Retrieve from https://dl-acm-org.ezproxy.une.edu.au/doi/abs/10.1145/3318396.3318419

[2] Herdlein, R., Riefler, L., and Mrowka, K. 2013. "An Integrative Literature Review of Student Affairs Competencies: A Meta-Analysis," Journal of Student Affairs Research and Practice (50:3), pp. 250-269. doi: 10.1515/jsarp-2013-0019

[3] Litster, K., and Roberts, J. 2011. "The Self-Concepts and Perceived Competencies of Gifted and Non-Gifted Students: A Meta-Analysis," Journal of Research in Special Educational Needs (11:2), pp. 130-140. doi: 10.1111/j.1471-3802.2010.01166.x

[4] Aishwarya, KL., Amuthan, F., Anjum, U., Bindu, K., and Iranna, SA. 2021, Aug. "Prediction of Student's Performance Based on Machine Learning, " in Journal of Research Proceedings (1:2), pp. 226-232.

[5] Bharara, S, Sabitha, S & Bansal, A 2018, 'Application of learning analytics using clustering data Mining for Students' disposition analysis', Education and Information Technologies, vol. 23, no. 2, pp. 957-984.

[6] Anderson, J. R., and Gluck, K. 2001. "What Role Do Cognitive Architectures Play in Intelligent Tutoring Systems," Cognition & Instruction: Twenty-five years of progress), pp. 227-262.

[7] Lewis, M., Milson, R., and Anderson, J. 1988. "Designing an Intelligent Authoring System for High School Mathematics Icai: The Teacher's Apprentice Project," Artificial intelligence and instruction: Applications and methods. New York: Addison-Wesley).

[8] Deejring, K. 2014. "The Design of Web-Based Learning Model Using Collaborative Learning Techniques and a Scaffolding System to Enhance Learners' Competency in Higher Education," Procedia-Social and Behavioral Sciences (116), pp. 436-441. doi: 10.1016/j.sbspro.2014.01.236

[9] Johnson, D. W., and Johnson, R. T. 2004. "Cooperation and the Use of Technology,").

[10] Anaya, A. R., and Boticario, J. G. 2011. "Application of Machine Learning Techniques to Analyse Student Interactions and Improve the Collaboration Process," Expert Systems with Applications (38:2), pp. 1171-1181. doi: 10.1016/j.eswa.2010.05.010

[11] Weimer, M. 2010. "Machine Teaching--a Machine Learning Approach to Technology Enhanced Learning." Technische Universität.

[12] Ren, Y., Zhang, L., and Suganthan, P. N. 2016. "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions," IEEE Computational Intelligence Magazine (11:1), pp. 41-53. doi: 10.1109/MCI.2015.2471235

[13] Zhang, L., and Suganthan, P. N. 2017. "Benchmarking Ensemble Classifiers with Novel Co-Trained Kernal Ridge Regression and Random Vector Functional Link Ensembles [Research Frontier]," IEEE Computational Intelligence Magazine (12:4), pp. 61-72. doi: 10.1109/MCI.2017.2742867

[14] Bryll, R., Gutierrez-Osuna, R., and Quek, F. 2003. "Attribute Bagging: Improving Accuracy of Classifier Ensembles by Using Random Feature Subsets," Pattern Recognition (36:6), pp. 1291-1302.

[15] Freund, Y., and Schapire, R. E. 1996. "Experiments with a New Boosting Algorithm," International Conference on Machine Learning, pp. 148-156. doi: 10.5555/3091696.3091715

[16] Amrieh, E.A., Hamtini, T. and Aljarah, I., 2016. "Mining educational data to predict student's academic performance using ensemble methods," International Journal of Database Theory and Application, 9(8), pp.119-136. doi: 10.14257/ijdta.2016.9.8.13

[17] Kotsiantis, S., Pierrakeas, C. and Pintelas, P. 2004. "Predicting Students' Performance in Distance Learning Using Machine Learning Techniques," Applied Artificial Intelligence (18:5), pp. 411-426. doi: 10.1080/08839510490442058

[18] Livieris, I.E., Drakopoulou, K. and Pintelas, P. 2012. "Predicting students' performance using artificial neural networks," in Proceedings of the 8th Pan-Hellenic Conference of Information and Communication Technology in Education. Doi: 10.14257/ijhit.2015.8.2.20

[19] Buenaño-Fernández, D. Gil, D. and Lujan-Mora, S. 2019. "Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study," Sustainability (11:10), pp. 1-18. doi: doi.org/10.3390/su11102833

[20] Pojon, M. 2017. "Using machine learning to predict student performance," MS thesis, University of Tampere, Faculty of Natural Sciences, 35.

[21] Khan, I., Al Sadiri, A., Ahmad, A. R. and Jabeur, N. 2019. "Tracking student performance in introductory programming by means of machine learning," in 4th MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1-6. doi: 10.1109/ICBDSC.2019.8645608

[22] Domingos, P. and M. Pazzani. 1997. "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning (29), pp. 103-130. doi: 10.1023/A:1007413511361

[23] Aha, D.W., Kibler, D. and Albert, M.K. 1991. "Instance-Based Learning Algorithms," Machine Learning (6), pp. 37-66.

[24] Simpson, O. 2006. "Predicting student success in open and distance learning," The Journal of Open, Distance and e-Learning (21:2), pp. 125-138. doi: 10.1080/02680510600713110

[25] Pang, Y., Judd, N., O'Brien, J. and Ben-Avie, M. 2017. "Predicting students' graduation outcomes through support vector machines," Frontiers in Education (FIE) Conference, pp.1-7. doi: 10.1109/FIE.2017.8190666

[26] Cortez, P. and Silva, A. 2008. "Using Data Mining to Predict Secondary School Student Performance," in 15th European Concurrent Engineering Conference, pp. 5-12.

[27] Alqurashi, T. and Wang, W. 2019). "Clustering ensemble method," International Journal of Machine Learning and Cybernetics (10:6), pp. 1227-1246.

[28] Jan, Z., Munoz, J. and Ali, Asim. 2020. " A novel method for creating an optimized ensemble classifier by introducing cluster size reduction and diversity," IEEE Transactions on Knowledge and Data Engineering, pp. 1-11. doi: 10.1109/TKDE.2020.3025173

[29] Asafuddoula, M., Verma, B. and Zhang, M. 2017, May. "An incremental ensemble classifier learning by means of a rule-based accuracy and diversity comparison," In 2017 International Joint Conference on Neural Networks (IJCNN), pp. 1924-1931. doi: 10.1109/IJCNN.2017.7966086