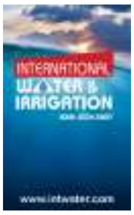


International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

Predicting Academic Success of Students Through Educational Data Mining

Fei Wang*

*Lecturer, Department of Literature, Art and Media, Tongling University, Tongling province, 244061, China

Abstract: This study uses four well-known machine learning methods—Lassis, Random Forest, AdaBoost, and Support Vector Regression—that are good at making predictions. It does this to look into the accuracy of different models and groups of variables. For this study, the China Education Panel Survey statistics were used. This survey took place between 2013 and 2015. The level of effort put in by each group has different effects on academic performance. Parents' strict expectations are the most important individual predictor of academic performance. School effort has a bigger effect on academic achievement than parental and student effort when different social background factors are taken into account. There are also big differences between male and female junior high school students in China, with school effort having a bigger effect on academic achievement. The results have big effects on strategies that are meant to get students to work harder in school. They show that treatments that are tailored to each gender are needed to help all kids do better in school.

Keyword; Datasets and Experiment by Paulo Cortez

1. Introduction

A lot of different parts of society have been changed by artificial intelligence (AI), such as the removal of basic factory work and the personalisation of ads based on user data. But these new technologies aren't often built into the learning management systems that are used in higher education right now [1, 2, 3].

Higher education focused on personalising, monitoring, and evaluating the learning process as COVID-19 came to an end [4,5]. Personalised education, rapid feedback, and focused practice are all opportunities that are being made available to students via the utilisation of these AI technologies [8].

Through the use of personalised learning that is customised to the specific requirements of each student, artificial intelligence enables pupils to study more efficiently and enhance their academic performance [11].

When students can customise their learning, they can find tools and tasks that fit their specific needs and preferences [14]. For people who learn best by doing, engaging activities and models may be helpful. For people who learn best by reading, including books, scientific journals, and technical studies [15], more in-depth academic materials are available.

When education is personalised, it enables us to uncover unique traits and group students, which in turn leads to improved student performance [18]. It is necessary to carefully plan and provide proper training for educational workers in order to successfully deploy artificial intelligence in higher education. Adaptive learning systems modify the degree of difficulty of assignments and tests according to the unique requirements and skills of every learner. This offers a customised learning experience and enables teachers to precisely evaluate each student's learning progress [6]. Adaptive learning systems are developed with the use of AI technologies.

It is possible to find kids who might not be able to stay in school, guess their grades, and make teaching methods more accurate with the help of these studies. A model for a college that can see into the future was made by Phan, De Caigny, and Coussement [23]. The outputs of this model are used by the authorities in order to make judgements, anticipate, and regulate the behaviours of students who drop out of school.

According to the findings of their investigation, Jiao [25] shown that a prediction approach that is based on evolutionary computation is a suitable instrument for evaluating the learning performance of students who are enrolled in online courses. In addition, the programming model demonstrates satisfactory performance in terms of prediction, which suggests that the most important factors in academic achievement are the acquisition of information, participation in class, and summative performance.

The model that was shown was a neural network model that could make correct predictions. There is still a gap, though: the study that is being done right now rarely looks at how students' learning changes over time or gives them feedback in real time on basic learning. To get the most out of AI in education, researchers should focus on making it easy to give quick comments and helpful knowledge that can be used right away. This will help bridge the gap between education interventions that are successful and forecasts made by AI.

These ideas, which come from identity danger and security theory, affect how well students who belong to groups that are overlooked or poorly portrayed do in school. Threats to women's identities and unfavourable preconceptions about their capabilities and potential in these sectors are commonplace for women. Empirical research, on the other hand, demonstrate the contrary. In college topics like computer science, statistics, and math, Bowman's study showed that female students did much better than male students.

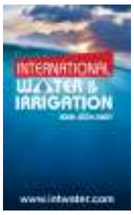
The expectancy-value theory of motivation developed by Eccles and Wigfield examines the ways in which the expectations and values that students have for a particular academic activity or assignment influence their level of motivation and performance. There is a connection between age and this idea since age may have an effect on the expectations and values that students have for their education. In this way, Portillo emphasised the importance of diversity in higher education.

The academic engagement hypothesis developed by Fredricks, Blumenfeld, and Paris places an emphasis on the active involvement and engagement of students in their educational experiences and in the classroom. Dhiman's research found that students' grades went down when they used cell phones and social networking sites for thirty minutes to three hours a day.

David Ausubel came up with the idea of useful learning, which Neroni talked about. This idea has something to do with people who use virtual rooms, digital books, and libraries for school. A study by De Jesús Araiza and García found that college students do better in their classes when they use technology tools like Red Canvas and Learning Management System (LMS) systems.

In reference to the research conducted by Sinchigalo and Arroba, who placed an emphasis on taking a more proactive and efficient approach to the academic growth of their students, we are developing a higher education system that is more individualised, efficient, and focused on the needs of the students.

The study will use a type of research called correlational–predictive research. This week's main task is to look at information about college students, especially when it comes to building and trying a model that can guess how well students will do in school. It is important to choose variables for a study that are useful, valuable, and backed by the academic literature. This is because the variables chosen need to be based on good theory. It is planned to pick variables on the basis of theoretical underpinnings that are supported by academic literature and theories that are pertinent to the study. When compared to traditional statistical methods like correlation and linear regression, machine learning has been shown to have a much higher level of accuracy (Table 1). For example, a machine learning algorithm is preferable than standard models when it comes to the analysis of significant amounts of internet data. This is because it permits relatively quick prediction while maintaining a high level of accuracy and big datasets.



International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

Table 1: ML Vs. Correlation and Linear Regression.

	Classical Statistical Method	ML Method	Reference
Rationale	Necessary for understanding the relationship between academic performance and relevant factors (e.g., crime rate and population density)	Prediction of academic performance by ML algorithms	(Bujang et al. 2021; Chang et al. 2020; Lykourentzou et al. 2009; Mduma et al. 2019; Papernot et al. 2017; Paulick et al. 2013; Sara et al. 2015)
Methods	The use of programs such as Mplus to identify relationships between academic performance and relevant factors; calculation based on the relationships	Prediction via 'black box' models without consideration for relationships	(Bujang et al. 2021; Chang et al. 2020; Lykourentzou et al. 2009; Mduma et al. 2019; Papernot et al. 2017; Paulick et al. 2013; Sara et al. 2015)
Accuracy	Existing relationships and assumptions	Quality and quantity of data	(Al-Jarrah et al. 2015; Ciolacu et al. 2017; Sekeroglu et al. 2019)
Advantages	Matured methods with clear processes	Rapid and convenient prediction for reasonable results	(Ciolacu et al. 2017; Sekeroglu et al. 2019)
Limitations	Sample selection bias	The 'black swan' effect	(Batrouni et al. 2018; Lorey et al. 2011)

2. Experiment And Analysis of Results

Paulo Cortez's earlier work [6] serves as a foundation for this study, which takes use of the same datasets.

2.1 Paulo Cortez Experiment and Datasets

The datasets for this investigation came from two Portuguese secondary schools. The Portuguese performance dataset comprises 649 student data, whereas the mathematics performance dataset has 395. There are 33 traits in two sets of data: 30 personal datapoints like school, age, gender, family education, and three grades for each of the three semesters. To set a standard, Paulo Cortez made a stupid guess. Based on facts about students, NN, SVM, and Random Forest (RF) are used to guess how well they will do. Table 1 shows how well the study's predictions were made [6], but it doesn't show the end grade.

In Paulo's study, a sorting technique is used to guess how well students will do. Paulo's study shows that a pass/fail method works better than a 5-level system for giving grades. The math dataset makes pass/fail grade predictions that are 21.7% more accurate on average, while the Portuguese dataset makes predictions that are 22.1% more accurate.

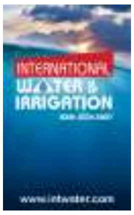
2.2 Regression Approach

The performance of students is represented as a number between 1 and 20 in the datasets that Paulo has created. The 5-level marking method then splits student work into five separate levels, as shown in Table 2. This makes the method much more private than it was before.

Table 2: Grading System With 5 Levels.

Grade	0~9	10~11	12~13	14~15	16~20
Level	1	2	3	4	5

The results of the performance that are presented in Table 1 indicate that, out of the four different DM approaches, SVM performs the lowest on the 5-level grading, whereas RF performs better. It is presumed that the regression technique is superior than the classification strategy when it comes to forecasting student performance. This is because RF is helpful for regression tasks in addition to classification activities. Also, using a 20-point scale to predict student success might give the truest picture of how students are doing on a steadier range. In order to compare the differences between a regression strategy and a classification strategy, this study uses the same DM methods as Paulo's research and a regression strategy. The study's theory is that a regression strategy works better than a classification strategy at predicting how well students will do in school.



International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

In the project being done for this study, a twenty-point grading system is used to train both NNs and RFs. It doesn't take training for the NNs and RFs to figure out how likely it is that a student "belongs to" a grade. They can guess a student's success value right away. Back in the day, NNs and RFs were not taught this way. This is the loss function that is used when both RFs and neural networks are being taught. The NNs are related to every layer. All of the RFs in this trial were made up of 1,000 estimators. Support Vector Machines (SVM) are no longer used for regression problems. As the Radial Basis Function (RBF), the Support Vector Regression (SVR) kernel function was chosen for this test. This feature projects information about how well students are doing onto a hyper-panel, which lets the SVR use the information to figure out the loss. During the grading process, each model makes a float number by guessing how well the students will do. Next, the closest integer from 1 to 20 to the float number is turned into a float number. Next, as shown in Table 3, the predicted grade is turned into a system with five levels of grade. The correctness of the estimate is then found by translating the marks from the five levels. One third of the dataset's total data is being used as test data in this study's experiment, and the other 70% is being used as training data. Table 3 shows the outcomes of the test.

Table 3: Regression Prediction Accuracy

	Mathematics			Portuguese		
	SVR	NN	RF	SVR	NN	RF
Regression	81.5%	73.4%	78.2%	75.9%	79.5%	72.8%
Classification	60.3%	59.6%	72.4%	65.1%	64.5%	73.5%
Improvement [1]	+21.9%	+13.1%	+5.8%	+11.4%	+14.4%	−0.7%

According to the results of the trial, SVR and NN that use the regression method do 15.2% better than those that use the classification method. The SVR for predicting maths achievement was more accurate, with a score of 21.9%. This shows how useful the regression method is for predicting how well children will do in school. There is some difference in how well RF works compared to SVR and NN, which could mean that an error happened during the experiment.

2.3 Synthetic Data Creation and Use

The outcomes in Tables 1 and 3 make it very clear that neural networks (NNs) do not as well on the mathematics dataset as they do on the Portuguese dataset. No matter if the regression or classification method is used, this is always the case. Also, neural networks like big training files because they are less likely to have been made by humans.

The dataset this study used to look at how well students could do in maths is smaller than the Portuguese dataset. There isn't enough information about how well kids do in math, so neural networks probably don't do well on the math dataset. This is because there are more users in the Portuguese collection. It is still very important to make estimates more accurate with smaller datasets.

Generative Adversarial Networks (GANs) are the names given to the adversarial networks that Goodfellow has been employing to create pictures for his study since 2014. [18] Goodfellow has been doing this research consistently since 2014. After GANs proved to be successful, many researchers used pictures made by AI to make models work better on smaller datasets [19,20]. Some computer programmes called GANs were chosen to help with this study's information on student maths performance by creating data on student performance.

The student data that was used for this study includes a total of 33 attributes, each of which is associated with a distinct value domain. The text values that are used by some of the features include things like gender, school names, and so on. This results in the student data being very confidential. Converting text data to a series of zeros and ones via the use of one-hot encoding is a typical approach to dealing with text data. On the other hand, this makes a lot more features in the data; after one-hot encoding, there will be 59 features, and most of the data in

International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

those features is zero, which doesn't carry much information. The data in Table 4 shows that GANs that were taught with data that had a single hot spot recorded can't make new data that correctly covers the original data's data distribution. Many things about them are the same, like their ages, Medu values, and Fedu values. The selected values are very similar to each other.After processing, the original outputs of GANs are changed so that they are in the value domain. This means that the GANs' first results were probably a lot worse. This is because one-hot encoded data doesn't work with GANs, as the test results show.

Table 4: A One-Hot Encoding GAN Sample

Age	Medu	Fedu	Absences	G1	G2	Traveltime	Studytime	Faliures	Famrel
15	1	0	1	1	0	1	1	1	1
15	1	0	1	1	0	1	1	1	1
15	1	0	0	0	1	1	1	1	1
15	1	0	1	1	0	1	1	1	1
15	1	0	1	1	0	1	1	1	1
15	1	0	1	0	1	1	1	1	1
15	1	0	0	0	1	1	1	1	1
15	1	0	1	1	0	1	1	1	1
15	1	0	1	1	0	1	1	1	1
15	1	0	0	0	1	1	1	1	1

An encoder and a decoder have been added to help the GANs. In this study, the encoder and decoder are taught with a three-layer neural network. Each layer has 59 dimensions: an input layer, a secret layer with 15 dimensions, and an output layer with 59 dimensions. You choose the weights between the secret layer and the input layer to set up the encoder. It is up to you to choose the weights that go between the secret layer and the output layer. This part is used to set off events in the encoder's output layer, which is secret. The output of encoding is scaled down to 0^1 by this function. To get the best results when encoding and decoding data, the whole student academic performance collection is used to teach both the encoder and the decoder. The eleven learned autoencoders are used to choose the one that works best (recovering 75.3% of the data after encoding) and insert it into the data. This structure is being talked about has a problem where the decoded data aren't always the same as the input. The auto-encoder is broken, which is why this trouble is happening. In each case, the encoder and decoder make a hyperspace out of the data dimension. In this case, the vector is denser. In this way, the problem that is caused by the very uneven spread of student data is lessened.

This lets us train and test the GANs. After the one-hot encoding process is done, this is what is done. The 15-dimensional vectors that were collected are used to teach the GANs in the next step. Following the training of the GANs, the output vectors from the generator are converted into one-hot vectors by the decoder. The one-hot vectors are then finally parsed into the standard student data format, as shown in Figure 1.

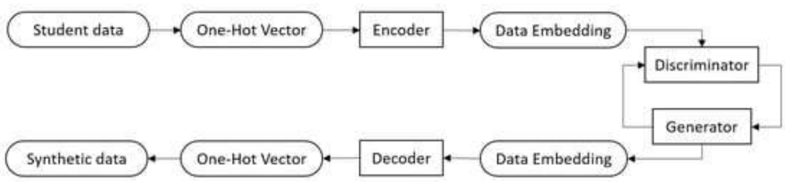
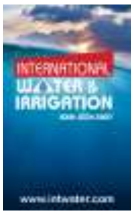


Figure 1: Workflow Data Creation

A generator and a discriminator are the two components that make up a GAN system. The data that is created by the generator is used to teach the discriminator to differentiate between them. During training, the generator is



taught to generate data that are as accurate as possible in order to deceive the discriminator. On the other hand, the generator will have the least loss—0—because the discriminator will think that all produced data is real data.

An illustration of the loss function for the discriminator is presented in the following example:

$$Loss_D = -\frac{1}{n}\sum_{i=1}^{n}\left[lnD\left(x^{(i)}\right) + ln\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right] = BCE(D(x), 1) + BCE(D(G(z), 0)) \tag{1}$$

Presented below is an illustration of the loss function for the generator:

$$Loss_D = -\frac{1}{n}\sum_{i=1}^{n} ln\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) = -BCE(D(G(z), 0)) \tag{2}$$

The variables in equations (1) and (2) are the collected data (x), the generator's input (z), the generated data (G(z)), the disambiguation result (D(a) for input (a), and the batch size (n). BCE is an acronym for binary cross entropy loss.

There are a lot of GANs being trained at the same time as the trial, and some of them don't change as planned. Some information from a generator that was left behind is shown in Table 5. The data that were picked show that this maker creates data whose feature values are alike. These created data can coincide with a particular point in the distribution of the actual data, making it impossible for the discriminator to differentiate between them.

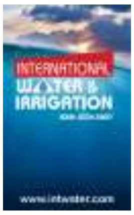
Table 5: Sample Abandoned GAN Data

Age	Medu	Fedu	Absences	G1	G2	Traveltime	Studytime	Faliures	Famrel
18	0	2	5	14	12	1	1	0	5
18	0	4	8	17	10	1	1	0	5
16	1	4	9	16	10	1	1	0	5
18	0	4	7	17	10	1	1	0	5
18	0	4	4	17	10	1	1	0	5
19	0	4	0	18	12	1	1	0	5
18	0	4	2	17	11	1	1	0	5
19	0	4	7	16	11	1	1	0	5
18	0	3	2	18	12	1	1	0	5
19	0	4	1	17	11	1	1	0	5

In the study written by Goodfellow, the advancement of GANs in picture production from 2014 to 2017 is just represented with an example during the course of the research period. When it comes to this investigation, the generator is selected according to the following guidelines:

For starters, the GANs ought to have a seamless conversion. It is seen in Figure 2 that the GANs that were abandoned undergo a loss shift that is quite rare. Those GANs that were selected for their smoother loss changes are shown in Figure 3, which depicts the converting process.

The second thing that the GANs need to do is make more spread-out data. While the GANs are being taught, it's possible that the generator will make data that is mostly in a certain range of how the real data is spread out. The discriminator is unable to account for this problem even if the generated data are correct and very comparable to one another. This difficulty may be solved by selecting a generator that creates data that is distributed over many locations.



International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

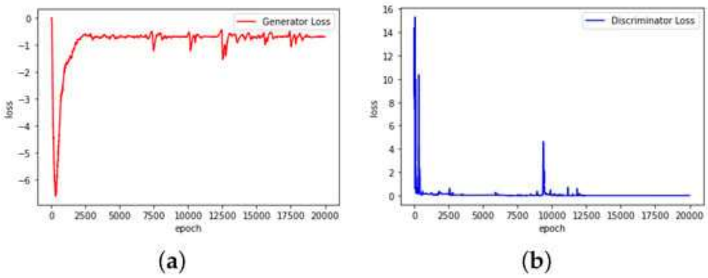


Figure 2: The Abandoned GAN Loss–Epoch Connection. Generator. B) Discriminator.

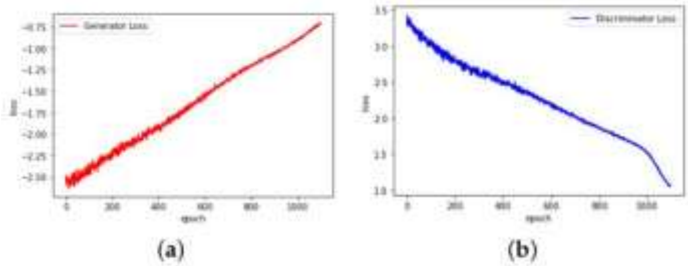


Figure 3: Generator And Discriminator of GAN Loss–Epoch Relationship.

Table 6 displays the examples of data that were created by GANs that were selected in accordance with the two principles described previously.

Table 6: Selected GAN Data.

Age	Medu	Fedu	Absences	G1	G2	Traveltime	Studytime	Faliures	Famrel
16	0	0	4	13	11	3	3	0	3
15	1	4	0	7	8	1	1	0	5
19	0	0	16	12	8	3	2	0	4
15	4	4	10	8	6	2	2	0	5
15	4	4	27	13	7	1	2	0	5
15	2	4	7	13	10	1	2	0	5
15	4	4	6	14	8	1	3	0	5
16	1	2	10	15	11	1	3	0	5
15	4	4	5	17	13	1	4	0	5
17	0	1	4	6	6	2	2	0	5

Using a regression method, the SVR, NN, and RF are taught again to guess how well students will do in maths, but this time on a dataset with fake data. This is done in order to test the effectiveness of the GANs. Seventy percent of the data in the training dataset was chosen at random from the mathematics performance dataset that was gathered. The training dataset comprises one hundred pieces of synthetic data. The remaining thirty percent of the data that was obtained is utilised as the test set. Table 7 displays the outcome of the investigation.

Table 7: Synthetic Data Regression Performance.

Methods	SVR	NN	RF
Performance	81.5%	75.9%	79.0%
Comparing to Classification Approach [1]	+21.9%	+15.6%	+6.6%
Comparing to Regression Approach [2]	+0%	+2.5%	+0.8%

An analysis of the findings that are shown in Table 7 reveals that while the accuracy of NN has improved by 2.5%, the accuracy of RF and SVR has remained essentially unchanged all throughout the process. Just because NN got a little more accurate, it looks like the data it made are a good match for how the real data are spread out. Table 6 shows a study of the created data. As the focused fake data are added, they only slightly even out the spread of the collected data. The neural network (NN) works a little better now. These traits are sometimes used as noise, but they also make it harder to understand the data.

3. Results, Discussion

In Table 8, you can see the overall evaluation scores for every model and dataset. Table 9 shows the overall rating scores for the dataset at the end of the first semester (S1). Table 10 shows the data for the set at the end of the second semester (S2). They are called Table 8, Table 9, and Table Ten, in that order.As part of the measures, the world F1 score for both the train set and the test set is given, along with the test set's balanced accuracy.

Table 8 Global assessment measures for each learning model generated using enrollment data (dataset S0). Average and SD are shown.

Table 8: Performance Comparison of Different Machine Learning Models on Dataset S0

Dataset S0			
ML Model	Train F1 Score	Test F1 Score	Test Balanced Accuracy
SMOTE + RF	0.779 ± 0.003	0.649 ± 0.007	0.654 ± 0.008
SVMSMOTE + RF	0.785 ± 0.004	0.649 ± 0.016	0.658 ± 0.018
BRF	0.631 ± 0.005	0.650 ± 0.025	0.643 ± 0.026
EE	0.620 ± 0.004	0.627 ± 0.026	0.621 ± 0.027
RB	0.598 ± 0.004	0.606 ± 0.020	0.600 ± 0.027

Table 9. Dataset S1 contains global assessment metrics for each learning model developed using first-semester data. Average and SD are shown.

Table 9: Performance Metrics of Machine Learning Models on Dataset S1

Dataset S1			
ML Model	Train F1 Score	Test F1 Score	Test Balanced Accuracy
SMOTE + RF	0.838 ± 0.003	0.709 ± 0.005	0.713 ± 0.005
SVMSMOTE + RF	0.841 ± 0.003	0.745 ± 0.014	0.749 ± 0.015
BRF	0.735 ± 0.003	0.744 ± 0.015	0.739 ± 0.015
EE	0.728 ± 0.005	0.729 ± 0.014	0.724 ± 0.015
RB	0.695 ± 0.004	0.702 ± 0.017	0.698 ± 0.017

Table 10. Dataset S2 contains global assessment metrics for each learning model developed using second-semester data. Average and SD are shown.

Table 10: Evaluation of Machine Learning Models on Dataset S2 Using Performance Metrics

Dataset S2			
ML Model	Train F1 Score	Test F1 Score	Test Balanced Accuracy
SMOTE + RF	0.888 ± 0.003	0.741 ± 0.006	0.746 ± 0.006
SVMSMOTE + RF	0.886 ± 0.002	0.741 ± 0.005	0.748 ± 0.005
BRF	0.727 ± 0.004	0.719 ± 0.018	0.711 ± 0.020
EE	0.727 ± 0.003	0.725 ± 0.015	0.714 ± 0.016
RB	0.650 ± 0.006	0.657 ± 0.025	0.654 ± 0.028

The models improve significantly globally from S0, where the highest F1 score is 0.650, to S1, when the best F1 score is 0.745. These models all demonstrate this increase. Considering that dataset S0 had no information on academic achievement after registration, you may have predicted this outcome. However, Dataset S1 already had some information on the student's college performance. This information was already available. The growth of these sorts of measurements becomes less evident when S2 models are included. The best F1 score for S2 is determined to be 0.741, which is quite near to the best F1 score for S1. Despite the fact that dataset S2 had more data on academic achievement, there was a more pronounced imbalance in the distribution of records throughout the classes. This makes things harder for the programmes that use machine learning.

During the training phase, SMOTE-based techniques include synthetic instances in the minority classes in order to guarantee that the datasets are proportional. When they are put through their paces, however, they are applied to datasets that are intrinsically uneven. As a consequence, the values of evaluation metrics are naturally equal for both training and testing.

Last but not least, the models SMOTE + RF, SVMSMOTE + RF, and BRF all result in superior outcomes compared to EE or RB. This remains true for each and every step that is discussed.

In addition to assessing global metrics, it is also important to analyse the performance of models for each class label. Tables 11, 12, and 8 give the F1 scores for each class, for each model, and for datasets S0, S1, and S2, respectively. These tables are arranged in sequential order.

Table 11. F1 score rating for each class and learning model based on enrollment data. Average and SD are shown.

Table 11: F1 Score Rating for Each Class And Learning Model Based On Enrollment Data

Dataset S0—F1 Score				
ML Model	Global	Graduate	Enrolled	Dropout
SMOTE + RF	0.649 ± 0.007	0.762 ± 0.008	0.493 ± 0.011	0.577 ± 0.009
SVMSMOTE + RF	0.649 ± 0.016	0.762 ± 0.022	0.454 ± 0.029	0.608 ± 0.026
BRF	0.650 ± 0.025	0.722 ± 0.028	0.511 ± 0.040	0.638 ± 0.075
EE	0.627 ± 0.026	0.706 ± 0.022	0.484 ± 0.037	0.606 ± 0.049
RB	0.606 ± 0.020	0.690 ± 0.024	0.433 ± 0.029	0.599 ± 0.029

Table 12. F1 score rating for each class and learning model based on first-semester data. Average and SD are shown.

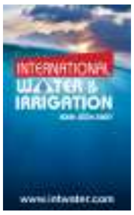


Table 12: F1 Score Rating for Each Class and Learning Model Based on First-Semester Data.

Dataset S1–F1 Score				
ML Model	Global	Graduate	Enrolled	Dropout
SMOTE + RF	0.709 ± 0.005	0.804 ± 0.005	0.560 ± 0.010	0.660 ± 0.008
SVMSMOTE + RF	0.745 ± 0.014	0.847 ± 0.010	0.555 ± 0.028	0.718 ± 0.030
BRF	0.744 ± 0.015	0.828 ± 0.016	0.582 ± 0.025	0.728 ± 0.026
EE	0.729 ± 0.014	0.814 ± 0.013	0.572 ± 0.025	0.705 ± 0.028
RB	0.702 ± 0.017	0.791 ± 0.018	0.524 ± 0.029	0.693 ± 0.030

Table 13. F1 score rating for each class and learning model based on second-semester data. Average and SD are shown.

Table 13: F1 Score Rating for Each Class and Learning Model Based on Second-Semester Data.

Dataset S2–F1 Score				
ML Model	Global	Graduate	Enrolled	Dropout
SMOTE + RF	0.741 ± 0.006	0.852 ± 0.005	0.616 ± 0.009	0.448 ± 0.022
SVMSMOTE + RF	0.741 ± 0.005	0.852 ± 0.005	0.615 ± 0.010	0.457 ± 0.017
BRF	0.719 ± 0.018	0.843 ± 0.016	0.541± 0.034	0.493 ± 0.055
EE	0.725 ± 0.015	0.841 ± 0.020	0.581± 0.031	0.455 ± 0.058
RB	0.657 ± 0.025	0.778 ± 0.028	0.499 ± 0.032	0.398 ± 0.109

The F1 marks for "Graduate," the class with the most students, are always higher than those for the other two classes. The classes with the next highest number are "Dropout" in dataset S1 and "Enrolled" in dataset S2. These classes also have the next highest scores.

The combination of SVMSMOTE and RF yields the greatest results for the "Graduate" class across all phases that are taken into consideration. BRF, on the other hand, is the one that achieves the greatest ratings for the "Enrolled" and "Dropout" classes for both S0 and S1. Because of this, the Balanced Random Forest models should be used to make the most accurate predictions for these types of groups.

Following the end of the second term, the "Dropout" class got the worst F1 scores out of all the classes that were part of the project. Not many records in dataset S2 are marked as "Dropout." The collection only has 398 records. There are 358 in the training set and 40 in the test set.

We can see the confusion matrices for each dataset and model in Figures 4, 5, and 6.

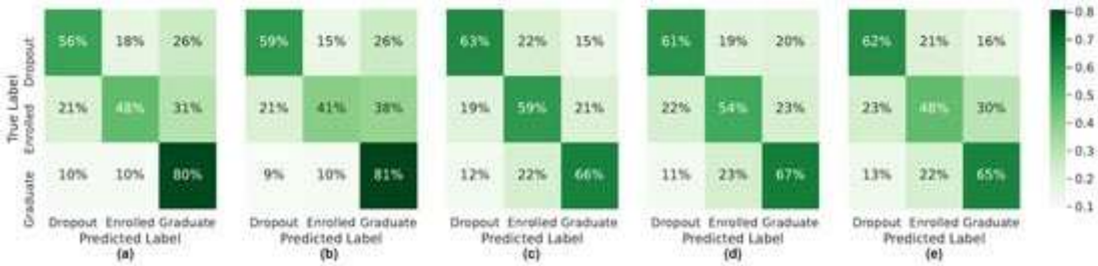


Figure 4: Set SVMSMOTE + RF, BRF, EE, And RB Confusion Matrices for S0 Models.

International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

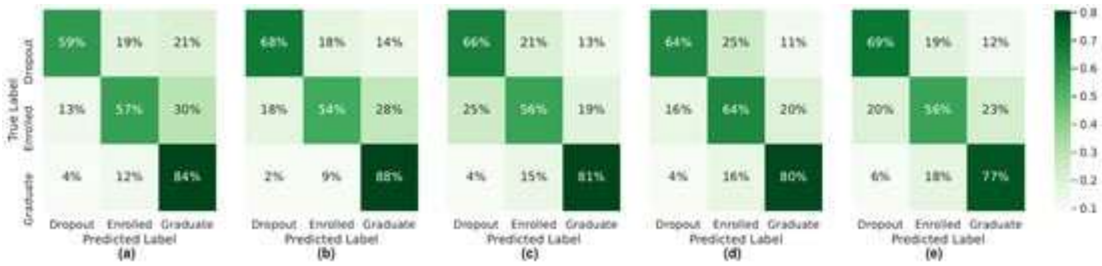


Figure 5: S1 Dataset Confusion Matrices: SMOTE, SVMSMOTE, BRF, EE, And RB.

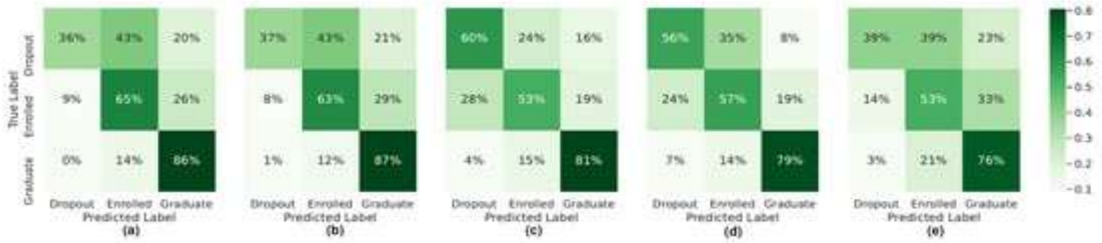


Figure 6: S2 Dataset Confusion Matrices: SMOTE, SVMSMOTE, BRF, EE, And RB.

Because of the low scores that are associated with the tiny number of students who are still tagged as "Dropout" in dataset S2, a question arises as to whether or not it would not be more appropriate to define the issue in this dataset as a binary one. In light of this subject, a second experiment was carried out using dataset S2. In this experiment, records that were originally named "Enrolled" and "Dropout" were changed to "Not Graduate." For the two-class classification problem, the data processing, model training, and evaluation steps were all done in the same way, as was already said. There is a presentation of the findings in Table 14.

Table 14. Evaluation of global and individual F1 scores for each learning model created using second-semester data from two courses. Average and SD are shown.

Table 14: Evaluation Of Global and Individual F1 Scores for Each Learning Model Created Using Second-Semester Data from Two Courses.

Dataset S2—F1 Score				
ML Model	Train	Test	Graduate	Not Graduate
SMOTE + RF	0.870 ±0.006	0.838 ± 0.013	0.866 ± 0.010	0.789 ± 0.016
SVMSMOTE + RF	0.870 ± 0.002	0.841 ± 0.015	0.869 ± 0.014	0.796 ± 0.020
BRF	0.828 ± 0.003	0.846 ± 0.021	0.871 ± 0.018	0.808 ± 0.026
EE	0.832 ± 0.004	0.840 ± 0.012	0.871 ± 0.011	0.792 ± 0.016
RB	0.813 ± 0.002	0.812 ± 0.014	0.846 ± 0.011	0.760 ± 0.018

On dataset S2, the Balanced Random Forest algorithm does the best job of classifying things into two groups. The best three-class score is 0.852 ± 0.005 for SVMSMOTE + RF, while the best class score is 0.871 ± 0.018 for BRF with the binary method. The class score for "Graduate" is higher than the best three-class score. There is also a lot more F1 score in the "Not Graduate" class.

These models were SMOTE + RF, SVMSMOTE + RF, and BRF. These show just how important the parts are.

International Water and Irrigation
Volume 44 Issue 2 (2025)
ISSN: 0334-5807
Pages 299 – 313

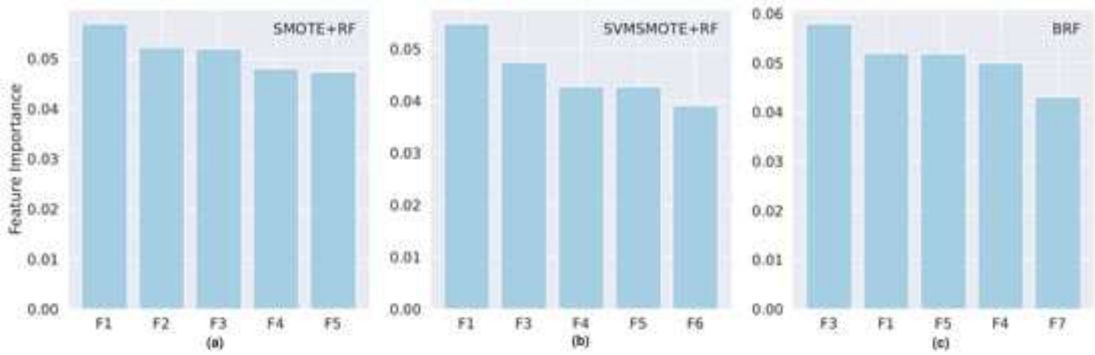


Figure 7: SMOTE, SVMSMOTE, And Balanced Random Forest Top the Five Most Essential Dataset S0 Model

Figure 7, SMOTE, SVMSMOTE, and Balanced Random Forest top the five most essential dataset S0 model attributes with the highest scores. F1: Age of enrollment F2: Current tuition; F3: Admission Grade; F4: Previous Habilitation Grade; F5: Displaced Student; F6: Degree; F7: National Exam Average.

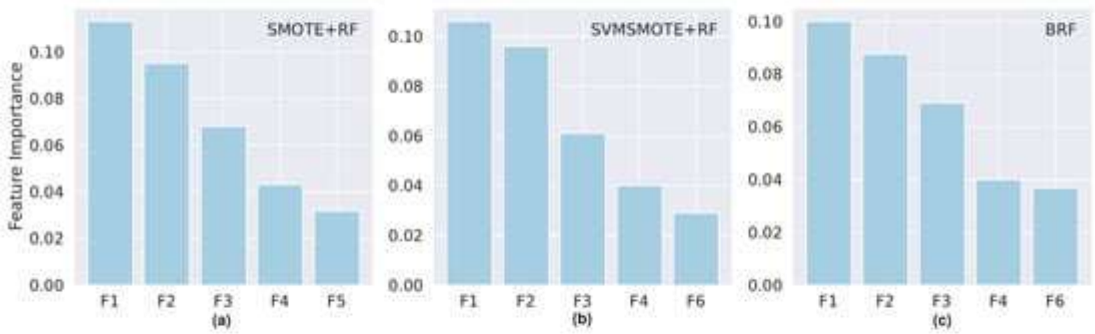


Figure 8: SMOTE, SVMSMOTE, And Balanced Random Forest Had the Greatest Five Dataset S1 Model Scores.

Figure 8. SMOTE, SVMSMOTE, and Balanced Random Forest had the greatest five dataset S1 model scores. F1: First-semester CUs authorized; F2: Average Grade; F3: Attendance; F4: Age upon enrollment; F5: Current tuition; Grade F6: Enter.

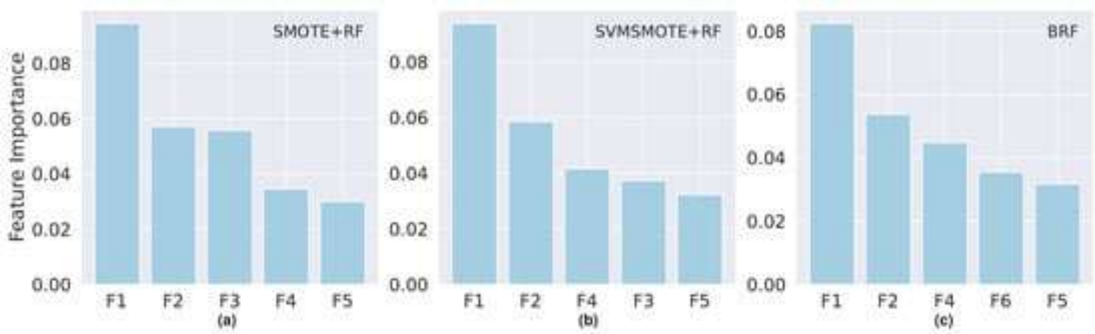
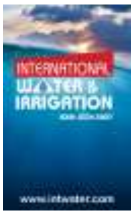


Figure 9: Top Five Most Essential Attributes for Dataset S1 Models with The Highest Scores

Figure 9. Top five most essential attributes for dataset S1 models with the highest scores: SMOTE, SVMSMOTE, and Balanced Random Forest. F1: Second semester CU approvals; F2: First semester approvals; F3: Current tuition expenses; F4: Second semester average grade; F5: Age at enrollment; F6: First semester average grade.



Among the many datasets, the most significant characteristics differ significantly from one another, whereas the differences across the models for the same dataset are quite slight. These are also things that all three types have in common. "Degree," "Average Grade at the National Exams," and "Up to date tuition fees" are in at least one of the models' top five lists of most important things. In all of these cases, the feature significance weights are very low, very close to the value of 0.05.

With regard to dataset S1, the three most significant characteristics that are shared by all of the models are the following: "Number of CUs approved (first semester)," "Average Grade (first semester)," along with "Number of CU attended (first semester)" . These traits will help a child do better in school once they start. In at least one model, the traits "Age at enrollment," "Tuition fees up to date," and "Admission Grade" are all in the top five, even though they have lower relative scores.

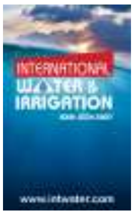
The most important traits in dataset S2 are those that have to do with academic progress. The "Number of CUs approved" in both the first and second semesters and the "Average Grade" in the second semester are two examples.

People who study the financial situation, the social and economic conditions, and the demographic conditions don't usually find them very important in the models. Unfortunately, this hurts the traits that are important for doing well in school after admission. Because dataset S0 doesn't have any information like this, these lines only work for datasets S1 and S2. Of all the social and economic factors that are thought to be important, "The only one that always makes it into the top five is "age at enrollment."The fact that all of the features in datasets S0 and S2 have feature significance ratings that are below 0.10 is something that should be taken into consideration at this point. There is a modest increase in the maximum feature significance score for the most significant feature in dataset S2, which is slightly over 0.10.

The findings that were obtained are comparable to those of other works that aim to forecast the academic performance or dropout rate of students at an early stage; nevertheless, our performance results are somewhat lower than those of other works. There was a dataset that was obtained from an engineering school. According to the findings of another research on performance prediction, the model attained an accuracy rate of 79% when it came to predicting the performance of students while they were applying for admission to a Computer Science degree programme. Binary classification problems and datasets derived from certain areas of study were used in each of these publications. This is in contrast to our multi-class issue, which was solved by employing a heterogeneous dataset. This dataset contains samples from a variety of different areas of study and schools, and it may be the reason why our models have somewhat lower performance ratings.

4. Conclusion

It is essential to both the execution of policies and the growth of society to be able to predict educational results. In order to provide light on the ways in which parental, school, and individual efforts may contribute to the formation and worsening of educational disparities, this research employs approaches from the field of machine learning. This study shows how important work is as a direct indicator of how well kids will do in school. If students are recognized for being hardworking, like by giving them more drive, determination, and patience when studying, their efforts may make up for the unequal educational resources they get from their families. This can help students maintain or improve their current academic standing as well as their future social position. It is also the primary conclusion of this study that efforts from both schools and parents are regarded as critical elements in enhancing educational achievements. This remains true whether the endeavours are examined via their distinct variables or seen as a pair of group variables. Thus, other methodologies may be considered in future study on social and educational disparity, especially when diverse effects are shown in individuals from different socioeconomic backgrounds. Encouragement of effort among groups who are marginally disadvantaged in social or academic contexts might be a more effective and efficient approach than waiting for their economic situation to improve.



Within the context of enhancing educational results, this research highlights the significant significance that school initiatives play. These policies include the "Quality Education" initiative, which invests in the training of teachers and educational materials, and the "Double Reduction" policy, which transfers academic responsibilities to families. This strategy would tackle the issue of inequalities in outcomes and efforts by enhancing the quality of education that is provided inside the classroom at every level of the educational process. The efficacy of this strategy is supported by the findings and findings. In light of this, educational institutions have to deliberately and intentionally plan and apply motivating tactics in order to provide a good and supportive learning environment for their classes of students.

References

[1] Salas-Pilco, S.Z.; Yang, Y. Artificial intelligence applications in Latin American higher education: A systematic review. Int. J. Educ. Technol. High. Educ. 2022, 19, 21.

[2] Khan, I.; Ahmad, A.R.; Jabeur, N.; Mahdi, M.N. An artificial intelligence approach to monitor student performance and devise preventive measures. Smart Learn. Environ. 2021, 8, 17.

[3] Bonilla-Jurado, D.; Zambrano, R.; Moncayo, H. Desarrolloprofesional continuo de docentes para actividadesenlosInstitutosTecnológicos: Unarevisiónliteraria. Rev. Científica UISRAEL 2023, 10, 27–39.

[4] Hwang, G.J.; Xie, H.; Wah, B.W.; Gašević, D. Vision, challenges, roles and research issues of Artificial Intelligence in Education. Comput. Educ. Artif. Intell. 2020, 1, 100001.

[5] Stadlman, M.; Salili, S.M.; Borgaonkar, A.D.; Miri, A.K. Artificial Intelligence Based Model for Prediction of Students' Performance: A Case Study of Synchronous Online Courses during the COVID-19 Pandemic. J. STEM Educ. 2022, 23, 39–46.

[6] Alqahtani, T.; Badreldin, H.A.; Alrashed, M.; Alshaya, A.I.; Alghamdi, S.S.; bin Saleh, K.; Alowais, S.A.; Alshaya, O.A.; Rahman, I.; Al Yami, M.S.; et al. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Res. Soc. Adm. Pharm. 2023, 19, 1236–1242.

[7] Mora, E.; Bonilla-Jurado, D.; Núñez, L.; Sarmiento, J. Inadaptabilidad de losDocentes al Manejo de PlataformasVirtuales: CasoEducarEcuador. Rev. Conrado 2018, 14, 83–90.

[8] Dai, Y.; Liu, A.; Lim, P.C. ReconceptualizingChatGPT and generative AI as a student-driven innovation in higher education. Procedia CIRP 2023, 119, 84–90.

[9] Al Ka'bi, A. Proposed artificial intelligence algorithm and deep learning techniques for development of higher education. Int. J. Intell. Netw. 2023, 4, 68–73.

[10] Gordillo-Tenorio, W.; Meléndez-Flores, C.; Sierra-Liñan, F.; Cabanillas-Carbonell, M. Information Technologies that Help Improve Academic Performance, A Review of the Literature. Int. J. Emerg. Technol. Learn. 2023, 18, 262–279.

[11] Hanaysha, J.R.; Shriedeh, F.B.; In'airat, M. Impact of classroom environment, teacher competency, information and communication technology resources, and university facilities on student engagement and academic performance. Int. J. Inf. Manag. Data Insights 2023, 3, 100188.

[12] Alfalah, A.A. Factors influencing students' adoption and use of mobile learning management systems (m-LMSs): A quantitative study of Saudi Arabia. Int. J. Inf. Manag. Data Insights 2023, 3, 100143.

[13] Melo, E.; Silva, I.; Costa, D.G.; Viegas, C.M.D.; Barros, T.M. On the Use of eXplainable Artificial Intelligence to Evaluate School Dropout. Educ. Sci. 2022, 12, 845.

[14] Barvinski, C.; Ferreira, G.; Machado, L.; Longhi, M.; Behar, P. Construction of a socio-affective profile model of students in a virtual learning environment. In Smart Education and e-Learning; Smart Innovation,

Systems and Technologies; Uskov, V.L., Howlett, R.J., Jain, L.C., Eds.; Springer Science and Business Media Deutschland GmbH: Porto Alegre, Brazil, 2019; pp. 159–168.

[15] Palacios, A.; Pascual, V.; Moreno-Mediavilla, D. The role of new technologies in STEM education. Bordon. Rev. De Pedagog. 2022, 74, 11–21.

[16] Tauber, Z.; Lacey, H.; Lichnovska, R.; Erdosova, B.; Zizka, R.; Sedy, J.; Cizkova, K. Students' preparedness, learning habits and the greatest difficulties in studying Histology in the digital era: A comparison between students of general and dental schools. Eur. J. Dent. Educ. 2021, 25, 371–376.

[17] Hinojo-Lucena, F.J.; Aznar-Díaz, I.; Cáceres-Reche, M.P.; Romero-Rodríguez, J.M. Artificial intelligence in higher education: A bibliometric study on its impact in the scientific literature. Educ. Sci. 2019, 9, 51.

[18] Díaz-Serrano, J.; Alfageme-González, M.B.; Cutanda-López, M.T. Interacción del rendimientoacadémico con losestilos de aprendizaje y de enseñanza. Rev. ElectrónicaInteruniv. De Form. Del Profr. 2022, 25, 145–160.

[19] Wang, Y. Artificial Intelligence Technologies in College English Translation Teaching. J. Psycholinguist. Res. 2023, 52, 1525–1544.

[20] Jeon, J.; Lee, S. Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Educ. Inf. Technol. 2023, 1–20.

[21] Lin, H. Influences of Artificial Intelligence in Education on Teaching Effectiveness. Int. J. Emerg. Technol. Learn. 2022, 17, 144–156.

[22] Rodríguez, M.d.L.F.; Suelves, D.M.; Méndez, V.G.; Mas, J.A.R.-L. Profesoradouniversitario. ¿Consumidor o productor de contenidosdigitaleseducativos? Rev. ElectrónicaInteruniv. De Form. Del Profr. 2023, 26, 13–25.

[23] Phan, M.; De Caigny, A.; Coussement, K. A decision support framework to incorporate textual data for early student dropout prediction in higher education. Decis. Support Syst. 2023, 168, 113940.

[24] Roselló, L.A.; Rodicio, C.I.F.; Suarez, D.C.R. Diferenciasen las percepciones del alumnadouniversitariosobreapoyosdocentessegúnmetodología, grado de estudios y edad. Rev. ElectrónicaInteruniv. De Form. Del Profr. 2023, 26, 193–205.

[25] Jiao, P.; Ouyang, F.; Zhang, Q.; Alavi, A.H. Artificial intelligence-enabled prediction model of student academic performance in online engineering education. Artif. Intell. Rev. 2022, 55, 6321–6344.