World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:19, No:07, 2025

Machine Learning Analysis of Student Success in Introductory Calculus Based Physics I Course

Authors : Chandra Prayaga, Aaron Wade, Lakshmi Prayaga, Gopi Shankar Mallu

Abstract : This paper presents the use of machine learning algorithms to predict the success of students in an introductory physics course. Data having 140 rows pertaining to the performance of two batches of students was used. The lack of sufficient data to train robust machine learning models was compensated for by generating synthetic data similar to the real data. CTGAN and CTGAN with Gaussian Copula (Gaussian) were used to generate synthetic data, with the real data as input. To check the similarity between the real data and each synthetic dataset, pair plots were made. The synthetic data was used to train machine learning models using the PyCaret package. For the CTGAN data, the Ada Boost Classifier (ADA) was found to be the ML model with the best fit, whereas the CTGAN with Gaussian Copula yielded Logistic Regression (LR) as the best model. Both models were then tested for accuracy with the real data. ROC-AUC analysis was performed for all the ten classes of the target variable (Grades A, A-, B+, B, B-, C+, C, C-, D, F). The ADA model with CTGAN data showed a mean AUC score of 0.4377, but the LR model with the Gaussian data showed a mean AUC score of 0.6149. ROC-AUC plots were obtained for each Grade value separately. The LR model with Gaussian data showed consistently better AUC scores compared to the ADA model with CTGAN data, except in two cases of the Grade value, C- and A-.

Keywords : machine learning, student success, physics course, grades, synthetic data, CTGAN, gaussian copula CTGAN
Conference Title : ICMLT 2025 : International Conference on Machine Learning Technologies
Conference Location : Stockholm, Sweden
Conference Dates : July 15-16, 2025