This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ACIT62805.2024.10877023, 2024 25th International Arab Conference on Information Technology (ACIT)

Evaluating the Impact of Multi-Layer Data on Machine Learning Classifiers for Predicting Student Academic Performance

Mohammad AlShaikh-Hasan
Computer Science Department
Brunel University London
London, UK
Mohammad.AlShaikhHasan@brunel.ac.uk

Gheorghita Ghinea
Computer Science Department
Brunel University London
London, UK
George.Ghinea@brunel.ac.uk

***Abstract*— It is important for educational institutions to be able to accurately assess students' performances to enable them to intervene and improve the learning process. This paper assesses the effectiveness of including multi-layer data in the performance of machine learning classifiers for student success prediction. By integrating student registration data with course-level Intended Learning Outcomes (ILOs), we compare several classifiers such as Logistic Regression, Random Forest, XGBoost, Support Vector Machines (SVM), and others. It is observed that integrating course-level ILOs data enhances classifier effectiveness in all metrics. For Random Forest, the accuracy improved to 0. 844 with the ILOs included compared to 0.770 using only registration data, but all others such as XGBoost and Logistic Regression demonstrate a significant improvement too. Therefore, the employment of multi layered data in addition to enhancing the predictive power of the models, gives institutions a holistic understanding of students' learning progression for immediate and focused interventions. These results reinforce our need to use rich data to improve the predictive power of academic achievement and student services.**

Keywords*— *Student Academic Performance, Machine learning, Prediction, Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), Decision Tree, AdaBoost, SVM, Multilayer Perceptron (MLP), XGBoost.

I. Introduction

Early identification of students who may be at the risk of poor academic performance is crucial as it allows educational institutions to address these issues before they escalate [1]. Most models use single layer data like demographics or exam results that may not necessarily represent a student's academic advancement or skills improvement [2]. As a result, these models do not always give the best understanding of factors that affect student success.

This limitation can be overcome through the use of multi-layer data, including registration details for the students and Course ILOs [3]. Course ILOs define the specific learning outcomes that students should achieve in a particular course and are more specific in describing in what ways students progress within a single course. By combining these data layers allows us to build models of both the immediacy of the learning process and the long-term effects of the learning process.

The purpose of this paper is to investigate the effects that multiple layers of data have on the performance of the machine learning classifiers in student performance prediction. In this work we demonstrate how these integrated data sources impact classifier performance (e.g., Logistic Regression, Random Forest, XGBoost, SVM), using Course ILOs and student registration information. The results indicate the need to incorporate more data layers to enhance the models predictive power, a finding that underlines the importance of rich datasets in the design of successful educational systems.

II. Related Works

The adoption of e-learning has transformed the educational scenario, with flexible, scalable, and diverse learning opportunities across different discipline areas. These platforms integrate high-level machine learning methodologies for higher predictions of probable student performance and, hence, much more effective educational interventions. In such a position, predictive models must identify the critical at-risk students and maximize learning outcomes with data-driven insights [4].

A study focused on the student dropout prediction in online university courses through the application of machine learning. The LMS data were collected, including student statistics data and records of interactions with the LMS. The four models used for early dropout were decision trees, random forest, support vector machines, and deep neural networks. The results showed that the random forest was the most efficient model with great accuracy and all the performance measures, which points at the use of machine learning in strengthening student support and retention in online learning environments [5].

Another successful prediction study was performed at the University of Jordan by [6]. They undertook a study to predict students' performance based on behavioural features. Moreover, traditional ensemble methods (such as Bagging, Adaboosting, and Random Forest) are used to predict the academic performance of students. The dataset was contained 480 students, whereby, the best accuracy of 79.1% is given by Adaboosting on Artificial Neural Network.

Mining educational textual data is also considered as a hot research domain that utilized machine-learning algorithms with educational data to figure out the performance of student learning outcomes. [7] highlight the critical effect of big data on higher education, training, learner experience and understanding, better academic research, better decision making and a planned response to varying general trends. Big data and analytical strategies were also recognized as common tools for achieving high-quality education benefits.

Neural networks and support vector machines can also be used to detect problems within the student learning process and improve learning performance by applying the results obtained from the assessment [8]. The authors highlight two main findings of the study, namely the effectiveness of early intervention as the key factor for students' success, and also the non-linearity of prediction logging Moodle results that indicate the applicability of new algorithms for analysing log data into many educational settings [8].

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ACIT62805.2024.10877023, 2024 25th International Arab Conference on Information Technology (ACIT)

A study reported in [9] explored the application of the ensemble method in predicting student performance in online learning. According to their study, they illustrated the computational methods that perfected the accuracy of classifying student success forecasts using a cluster prediction method. This strategy involved using more than one classifier for the whole Kalboard 360 program support dataset.

In another recent study, [10] developed a model using the Firefly Algorithm to time academic performance at the Saudi Electronic University in a blended learning environment. The results of their study showed that the model is capable of identifying the risk factors for low performance by students include attendance, midterm exams and assignment submission. This approach provides a new dimension in educational analytics, offering potential pathways for personalized interventions to enhance student outcomes.

In related work, [11] conducted a study in which they developed a model to predict students' performance and activities using machine learning algorithms. Their objectives were to develop a model for predicting overall performance (grade/engagement) of the students, and to understand the impact of the features in online learning platform on its student outcomes. The foundation of their study adopted a quantitative method in handling and processing of student's data. The outcome of the study also revealed that the Random Forest classifier proved to yield the high overall accuracy of 85% in predicting grades and 83% for the engagement of the students based on record on a set of attributes that involves students' profile and data on the interaction of the students within the online learning environment [11].

Another research endeavour [12] described the development of a new approach that uses a combination of unsupervised and supervised learning for the prediction of student outcomes regarding their study in higher education. Their case study was done among three students of the University of Thessaly in Greece, in the context of a Computer Science programme. First, the quantitative study analysed quantitative variables extracted from the survey data using the K-Means clustering algorithm, which it used to assign students to one of three clusters based on their educational factors. Subsequently, for each student cluster, predictive models involving supervised form of machine learning were used to determine the time required to complete a degree as well as student enrolment in education program. The approach described here was seen to predict well as based on the results several accurate case study predictions were made. The authors consider the proposed approach, based on the combination of unsupervised and supervised learning, applicable to the field of learning analytics and predicting students' outcomes in higher education settings [12].

III. Data Description

The study described in this paper utilizes two datasets to predict student academic performance by applying and evaluating various machine learning classifiers. These datasets offer different layers of student information, providing comprehensive insights into each student's academic journey.

A. Data-1 (Registration Data)

The dataset consists of registration academic information for students of IT college at the University of Petra, Jordan, for the 2022–2023 academic year, which includes details from the first and second semesters. This dataset consists of 952 records, each reflecting a unique student, and collects demographic and academic information about student progress. **Table I** outlines the key features of the registration data.

TABLE I. REGISTRATION-DATA KEY FEATURES

Feature	Description	Example
Anonymized StdID	Unique, anonymized identifier for each student.	2190141954851600
Major	The student's major	Computer Science
Nationality	The student's nationality	Jordanian
High school type	The type of high school education received	Scientific
High school country	The country where the student completed high school.	Jordan
Year of high school	The year of high school graduation	2020
High school score	The student's high school GPA or score	85.4
Enrollment year	The year the student enrolled in the university	2020
Hours registered	The total number of credit hours registered in the current semester	133
Hours completed	The total number of credit hours successfully completed	97
Date of Birth	The student's birthdate, used to calculate their current age	24-03-2002
Student GPA	The student's GPA in the current program of study	2.15

The dataset serves as an important base for predictive modeling in the study, by providing key demographic and academic information.

B. Data-2 (Intended Learning Outcomes Data)

This dataset records student performance relative to multiple Intended Learning Outcomes (ILOs) for IT college students at the University of Petra, Jordan, during the academic year 2022-2023, from both semesters. The records included in the dataset total 24,798 and provide specific evaluations of student performances concerning a range of learning objectives related to individual courses. All of the programs in this college use benchmarks for student outcomes that are measured according to international accreditations [13]. **Table II** outlines the key features of the ILOs data.

ILOs are specific and measurable goals which provide specific skills, knowledge, and competencies that students should acquire at the end of each course. In this study, each course in the IT college is designed with distinct ILOs based on international certification standards so as to reflect academic and industry expectation. The outcomes of these are across many learning domains including theoretical knowledge, practical skills and analytical abilities, all evaluated against specific evaluation criteria. This dataset allows us to have a precise measure of academic attainment in relation to these ILOs to help us to determine how well the students are learning, and in turn, where there are strengths and gaps in what the students are learning. By taking this approach, the study can then capture a more nuanced understanding of student progression than traditional course level assessment metrics.

These data sources came from the extended Moodle system in use at the IT collage of the University of Petra. Calculations for this dataset were made for every student in

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ACIT62805.2024.10877023, 2024 25th International Arab Conference on Information Technology (ACIT)

every course in accordance with international accreditation standards. The extended Moodle system gathers and evaluates the progress of each student toward specified course outcomes. By the end of each semester, the Moodle collects data from all courses and produces a flat table [3].

Importantly, the analytic data in this dataset is collected at the micro level rather than the total assessment level. For instance, instead of using a student's entire exam score to measure course outcomes, the system links specific exam questions to different course outcomes, providing a more granular and detailed analysis of student performance.

TABLE II. ILOS-DATA FEATURES

Feature	Description	Example
Anonymize StdID	Unique, anonymized identifier for each student.	2190141954851600
Year	The academic year during which the course was taken	2022
Semester	The semester (first or second) in which the course was taken	1
Course ID	Identifier for the course associated with the ILO	601111
CILO	The specific Course Intended Learning Outcome assessed	K1
Std Attainment	The student's attainment score for the specific ILO in the course	0.774

Since this dataset provides for a granular analysis of student performance, it allows for developing highly detailed understanding of which areas of learning objectives students have excelled at and continue to struggle with.

C. *Data Integration*

The Anonymized_StdID from both datasets were used to integrate the datasets. The integration makes it possible to go beyond traditional predictions by analyzing student performance holistically using registration data alongside ILOs to improve the depth and accuracy of prediction models.

IV. METHODOLOGY

This section outlines the methodology applied to assess the impact of multi-layer data on the performance of machine learning classifiers for predicting student academic performance. We conduct 2 experiments: experiment1 on registration data alone and experiment2 on combining registration data with the ILOs data.

The methodology as shown in **Fig1** below has several stages to predict the students' performance. It starts with the data preprocessing phase, where student registration data (Reg), and ILOs data are cleaned and organized. Next, feature selection is used to feature the most important variables affecting students outcomes and reduces the amount of work in the ML process. To counter imbalance in the class data set, the SMOTE (Synthetic Minority Over-sampling Technique) is used, which synthesizes samples in an under-representative class [14]. Thereafter, the hyperparameter of the models is optimised to enhance the models' performance. To the given data, several ML algorithms are applied. Using cross-validation, the models are trained and tested to ensure reliable performance. Finally, the models' effectiveness is evaluated using performance metrics like accuracy, precision, recall, and F1 score to decide which algorithm was performed well [16].

A. *Data Preprocessing:*

Several preprocessing steps were performed:

1) *Missing Values:* Numeric columns were imputed with the mean value to handle missing data.

2) *Feature Engineering:* A new feature, age, was created from the date of birth. In experiment2, ILO data were aggregated by student ID using summary statistics (mean, sum, max, and min).

3) *Categorical Encoding:* Categorical variables, such as major, and nationality, were encoded using label encoding.

4) *GPA Categorization:* GPA was binned into three classes—low (0–2.0), medium (2.0–3.0), and high (3.0–4.0).

B. *Feature Selection:*

Feature Selection helped us identify the top relevant predictors of student performance. In the Level-1 Experiment, features such as nationality, high school score, enrollment year, hours registered, hours completed and age are chosen. We also added some additional features in the Level 2 Experiment such as ILO statistics (mean, sum, max, and min) for additional depth in predictions. The combination of Logistic Regression and Recursive Feature Elimination (RFE) was used to select the most valuable features for performing feature selection. This careful selection process ensured that the subsequent ML models would have a refined set of features to work with, ultimately improving their predictive power.

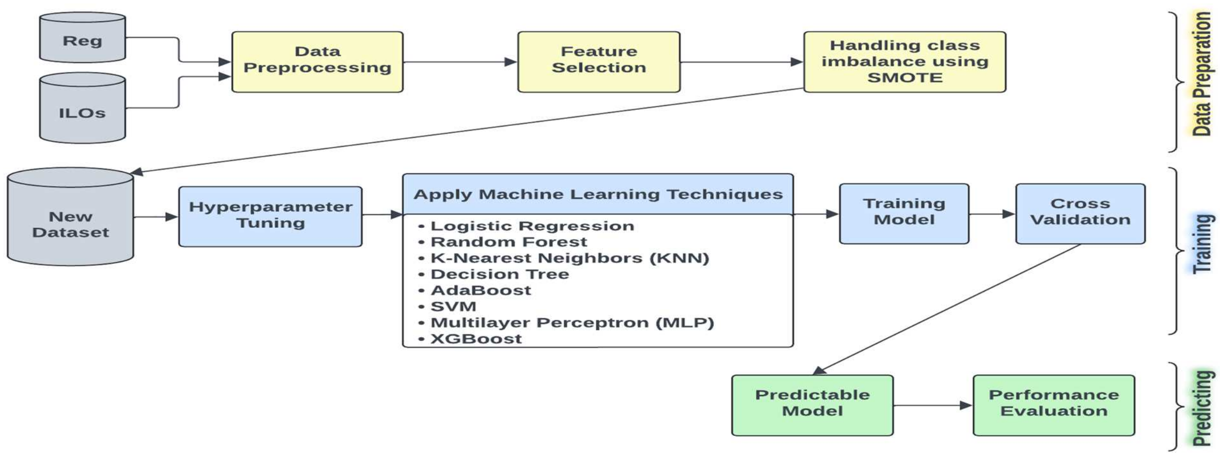

Fig. 1. Proposed methodology

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ACIT62805.2024.10877023, 2024 25th International Arab Conference on Information Technology (ACIT)

C. Machine Learning Classifiers:

Several machine learning classifiers were applied to both datasets. Each classifier was optimized through hyperparameter tuning and evaluated for its performance on both registration data alone (Experiment1) and combined registration and ILO data (Experiment2). The classifiers included:

*1) **Logistic Regression**:* A linear model used for multi-class classification.

*2) **Random Forest**:* An ensemble method that constructs multiple decision trees and aggregates their outputs.

*3) **K-Nearest Neighbors (KNN)**:* A non-parametric classifier that predicts the class based on the majority vote of the nearest neighbors.

*4) **Decision Tree:*** A tree-based model that splits the data based on feature values.

*5) **AdaBoost:*** An ensemble technique that combines weak classifiers to improve performance.

*6) **Support Vector Machine (SVM):*** A classifier that seeks to find the hyperplane that best separates the classes.

*7) **Multilayer Perceptron (MLP):*** A neural network-based model.

*8) **XGBoost:*** A gradient-boosting model optimized for high performance and fast execution.

D. Addressing Class Imbalance with SMOTE:

In order to resolve class imbalance in this dataset we used the Synthetic Minority Over-sampling Technique (SMOTE). An oversampling method SMOTE generates synthetic samples in the minority class by interpolating between existing minority samples and their closest neighbors [14]. This technique achieves a perfect balance between classifier distribution while generating new data instances along line segments between classes, i.e., introducing robust models without any duplication of existing instances.

SMOTE was applied after the data preprocessing stage. The technique was implemented using Python's imbalanced learn library and optimized parameters that balance the classes while not adding noise. To be able to generalize well across the different student performance categories, the classifiers needed to be able to learn well from representative training data across each class, so we applied SMOTE to ensure the training data had enough data from each class.

E. Hyperparameter Tuning:

For each classifier, hyperparameter tuning was conducted using GridSearchCV to optimize performance. The specific hyperparameters for each classifier were systematically adjusted to find the best model configuration. **Table III** outlines the key hyperparameters that were tuned for each machine learning algorithm:

TABLE III. HYPERPARAMETERS TUNED FOR EACH CLASSIFIER

Classifier	Tuned Hyperparameters
Logistic Regression	Regularization strength (C), Solver (solver)
Random Forest	Number of trees (n_estimators), Max depth (max_depth)
K-Nearest Neighbors (KNN)	Number of neighbors (n_neighbors), Distance metric (p), Weights (weights)
Decision Tree	Max depth (max_depth), Min samples split (min_samples_split)
AdaBoost	Number of estimators (n_estimators), Learning rate (learning_rate)
SVM	Regularization (C), Kernel (kernel), Gamma (gamma)
Multilayer Perceptron (MLP)	Hidden layers (hidden_layer_sizes), Activation function (activation), Learning rate (learning_rate_init)
XGBoost	Number of estimators (n_estimators), Learning rate (learning_rate), Max depth (max_depth)

F. Cross-Validation:

A 10-fold Stratified K-Fold cross validation scheme was used to prevent overfitting and generalize the model better, keeping class distribution in the training and test folds the same. This approach ensures a robust evaluation by checking the model on different subsets of data, and simultaneously prevents overfitting on any one partition of the dataset.

G. Evaluation Metrics:

The performance of each classifier was evaluated using:

*1) **Accuracy**:* The proportion of correctly classified instances *[16]*.

*2) **Recall (macro-averaged)**:* The ability to identify relevant instances across classes *[16]*.

*3) **Precision (macro-averaged)**:* The accuracy of predictions for each class *[16]*.

*4) **F1 Score (macro-averaged):*** The harmonic mean of precision and recall *[16]*.

H. Model Implementation:

Python was used to implement all models. scikit-learn, xgboost, and pandas were key libraries used for data manipulation and modeling. The SMOTE was applied to handle class imbalance [14].

V. RESULTS

This section presents the outcomes of evaluating machine learning classifiers on two experiments: The first one is Experiment1 (Exp1-Reg) based only on students' registration data and the second one is Experiment2 (Exp2-Reg-ILOs) supplemented with ILOs data. All measures of evaluation such as accuracy, recall, precision, and F1 score were considered while assessing both experiments.

A. Level-1 Experiment (Exp1-Reg) Results

Experiment1 used only registration data for IT college students. The following classifiers were applied, and their performance metrics are listed in **Table IV** below.

Analysis of Experiment1 Results:

- Random Forest and XGBoost were the top performers, with Random Forest achieving the highest accuracy (0.770) and F1 score (0.767).
- KNN and Decision Tree showed moderate results.
- Logistic Regression showed the lowest scores across all metrics, with MLP performing modestly, and SVM and AdaBoost showing relatively lower scores.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ACIT62805.2024.10877023, 2024 25th International Arab Conference on Information Technology (ACIT)

TABLE IV. PERFORMANCE OF CLASSIFIERS FOR EXPERIMENT1 (REG)

Classifier	Accuracy	Recall	Precision	F1 Score
Logistic Regression	0.589	0.589	0.596	0.584
Random Forest	0.770	0.770	0.770	0.767
K-Nearest Neighbors	0.685	0.685	0.686	0.678
Decision Tree	0.677	0.677	0.676	0.674
AdaBoost	0.626	0.626	0.627	0.621
Support Vector Machine	0.670	0.670	0.671	0.667
MLP Classifier	0.675	0.675	0.678	0.672
XGBoost	0.764	0.764	0.768	0.762

B. *Level-2 Experiment (Exp2-Reg-ILOs) Results*

Experiment2 incorporated both registration data and course-level ILOs. The following classifiers were applied, and their performance metrics are listed in **Table V** below:

TABLE V. PERFORMANCE OF CLASSIFIERS FOR EXPERIMENT2 (REG-ILOS)

Classifier	Accuracy	Recall	Precision	F1 Score
Logistic Regression	0.781	0.782	0.782	0.779
Random Forest	0.844	0.844	0.847	0.842
K-Nearest Neighbors	0.832	0.832	0.833	0.828
Decision Tree	0.768	0.768	0.771	0.767
AdaBoost	0.768	0.768	0.774	0.768
Support Vector Machine	0.837	0.837	0.837	0.834
MLP Classifier	0.787	0.787	0.793	0.787
XGBoost	0.845	0.846	0.849	0.843

Analysis of Experiment2 Results:

- Experiment2 showed significant improvements with the inclusion of ILO data.
- Random Forest and XGBoost were the top classifiers, with XGBoost achieving the highest accuracy (0.845) and F1 score (0.843), closely followed by Random Forest with an accuracy of 0.844 and F1 score of 0.842.
- Logistic Regression improved dramatically from Experiment1, indicating the benefit of incorporating ILO data for even simpler models.

The results show the value of utilizing additional details, like the ILOs, to improve the performance of the machine learning classifiers. Random Forest and XGBoost models showed the highest accuracy across both experiments and there was a notable increase when incorporating the ILO data. These findings are significant because they show that more granular and complex data are capable of generating higher predictive power and, therefore, more meaningful information in terms of students' performance and their needs.

VI. DISCUSSION

The study shows that the use of multi-layer data, especially ILOs, is useful in predicting student's academic performance in machine learning models. Experiment2 (Exp2-Reg-ILOs) which incorporated the registration data with the ILOs performed better than Experiment1 (Exp1-Reg) which only used registration data. The discussion addresses the following points:

A. *Effect of Multi-Layer Data on Classifier Performance*

The addition of multi-layer data improved the classifier performance in all metrics considered with relatively significant improvement. For instance, Random Forest accuracy rose from 0.770 (Experiment1) to 0.844 in Experiment2, and XGBoost also got a similar improvement from 0. 764 to 0. 845.

The enhancement of recall, precision and F1 score indicates that the multi-layer data allows for a better understanding of the student's progress and as a result, better predict their academic achievement. Logistic Regression, a simpler model, also benefited from ILO inclusion, with its accuracy rising from 0.589 to 0.781.

B. *Importance of ILO and Program Outcomes Data*

The specific ILO information offers more specific information on the students' achievement in terms of the learning outcomes. Such micro level analysis can be used in conjunction with macro indicators such as GPA, to aid classifiers to come up with a more accurate picture of a student's performance. Random Forest and XGBoost delivered the best results with this enriched data in which the potential connections between registration data and certain learning outcomes were utilized to provide more accurate and individualized predictions.

C. *Model Comparison and Implications for Educational Data Science*

Out of all the models, ensemble models such as Random Forest and XGBoost performed the best especially in Experiment2. These models are particularly ideal for data with many dimensions such as detailed ILOs and registration data. However, models like SVM and MLP were less accurate particularly in the multi-class classification based on GPA which showed that there is still room for improvement in the extraction of the full potential of the multi-layer data.

D. *Practical Implications for Educational Institutions*

The implications of the study are evident to institutions that seek to enhance on the models that predict student performance. Thus integrating multi-layer data institutions can develop better models that enable early interventions. This approach will help educators in the identification of the learners who might not be able to achieve the intended learning outcomes thus helping them to provide support to the affected learners at the right time. Such individualised interventions could result in increased student retention and better academic achievement.

E. *Limitations and Areas for Future Research*

1) Scope: This study uses data from a single academic year within one college. While this dataset includes a diverse range of student demographics, academic backgrounds, and detailed course-level Intended Learning Outcomes (ILOs), enhancing generalizability remains a priority. To address this,

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/ACIT62805.2024.10877023, 2024 25th International Arab Conference on Information Technology (ACIT)

robust k-fold cross-validation was applied to ensure the reliability of findings across various data partitions. Future work could expand the dataset by including data from multiple colleges or academic years, which would enable further exploration of the generalizability of these findings across different educational settings.

2) Additional Data Layers: It is important to note that other variables like student's engagement or interactive online activities could also improve the models..

3) Advanced Techniques: Possible future work could involve the use of more advanced models such as deep learning or reinforcement learning that may be better suited to model the interactions that are present in multi-layered educational datasets.

VII. Conclusion

This research highlights the value of incorporating multiple layers of data, including registration information and ILOs, into machine learning models for predicting student performance. The findings show that using ILOs improves the accuracy of the prediction models greatly over registration data especially with the use of ensembles such as Random Forest and XGBoost. For example, Random Forest improved from 0.770 in Experiment1 to 0.844 in Experiment2, while Logistic Regression improved from 0.589 to 0.781.

ILOs help models to evaluate the student's achievement and learning outcomes in more detail. This results in much better predictions, which enable institutions to provide specific support to students who may be in danger of failing.

These results confirm the real implications for educational organisations willing to enhance the performance of students using predictive analytics. This way, institutions can develop personalized learning strategies and offer timely interventions to students.

References

[1] O. Zawacki-Richter *et al*, "Systematic review of research on artificial intelligence applications in higher education–where are the educators?" *International Journal of Educational Technology in Higher Education,* vol. 16, *(1),* pp. 1–27, 2019. .

[2] Z. Shou *et al*, "Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach," *Applied Sciences,* vol. 14, *(6),* pp. 2522, 2024. .

[3] M. A. Arqoub *et al*, "Extending learning management system for learning analytics," in *2022 International Conference on Business Analytics for Technology and Security (ICBATS),* 2022, .

[4] A. F. Gkontzis *et al*, "An effective LA approach to predict student achievement," in *Proceedings of the 22nd Pan-Hellenic Conference on Informatics,* 2018, .

[5] H. S. Park and S. J. Yoo, "Early dropout prediction in online learning of university using machine learning," *JOIV: International Journal on Informatics Visualization,* vol. 5, *(4),* pp. 347–353, 2021. .

[6] E. A. Amrieh, T. Hamtini and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods," *International Journal of Database Theory and Application,* vol. 9, *(8),* pp. 119–136, 2016. .

[7] A. Klašnja-Milićević, M. Ivanović and Z. Budimac, "Data science in education: Big data and learning analytics," *Computer Applications in Engineering Education,* vol. 25, *(6),* pp. 1066–1078, 2017. .

[8] M. Ciolacu *et al*, "Education 4.0—Fostering student's performance with machine learning methods," in *2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME),* 2017, .

[9] F. Saleem *et al*, "Intelligent decision support system for predicting student's e-learning performance using ensemble machine learning," *Mathematics,* vol. 9, *(17),* pp. 2078, 2021. .

[10] N. N. Hamadneh *et al*, "Using artificial intelligence to predict students' academic performance in blended learning," *Sustainability,* vol. 14, *(18),* pp. 11642, 2022. .

[11] Y. T. Badal and R. K. Sungkur, "Predictive modelling and analytics of students' grades using machine learning algorithms," *Education and Information Technologies,* vol. 28, *(3),* pp. 3027–3057, 2023. .

[12] O. Iatrellis *et al*, "A two-phase machine learning approach for predicting student outcomes," *Education and Information Technologies,* vol. 26, pp. 69–88, 2021. .

[13] N. El-khalili *et al*, "Empowering learning analytics with business intelligence," in *2024 2nd International Conference on Cyber Resilience (ICCR),* 2024, .

[14] M. Al-Nashashibi *et al*, "A New Two-step Ensemble Learning Model for Improving Stress Prediction of Automobile Drivers," .

[15] M. Y. Al-Nashashibi, M. Alauthman and W. Hadi, "Machine learning for enhanced autism screening: A comparative evaluation of classification algorithms," in *2024 2nd International Conference on Cyber Resilience (ICCR),* 2024, .