

Supporting Student Success Through an Interpretable Early Warning System in a First-Year Statistics Unit

Bee Lam and Steven Richardson

Edith Cowan University, Australia

Abstract

Early identification of students experiencing academic difficulty is critical to supporting student success and retention in their first year of university. This study examines an interpretable learning analytics framework designed to support timely intervention in a first-year introductory statistics unit. Data from 566 students across three cohorts were used to develop and evaluate predictive models using assessment results collected during the teaching period and enrolment characteristics. The selected model demonstrated strong predictive performance and was validated on a subsequent cohort. To support practical implementation, predictions were translated into a colour-coded risk report that enables educators to quickly identify and prioritise students requiring academic support. SHapley Additive exPlanations (SHAP)-based explanations were incorporated to provide transparency and assist educators in understanding the factors contributing to individual risk classifications. Findings indicate that formative assessments are the strongest indicators of academic success, highlighting opportunities for proactive and targeted intervention. The framework demonstrates how interpretable learning analytics can support evidence-informed decision-making and student success in large first-year units.

Keywords: Student success; early intervention; learning analytics; first-year students; predictive analytics; explainable AI; higher education.

Introduction

Student success in the first year of university remains a key priority for higher education institutions. Students who experience academic difficulty during the early stages of their studies are more likely to disengage, withdraw, or fail to progress, highlighting the importance of timely identification and support (Kahu & Nelson, 2018; Kift et al., 2010). Within large first-year units, however, identifying students who may require additional support can be challenging, particularly when teaching staff are responsible for large cohorts and have limited visibility of emerging academic difficulties.

Learning analytics offers opportunities to support evidence-informed approaches to student success by using routinely collected educational data to identify students who may be at risk of poor academic outcomes (Brooks & Thompson, 2022; Namoun & Alshanqiti, 2021). Previous studies have demonstrated that academic performance can be predicted with reasonable accuracy using combinations of assessment data, demographic variables, enrolment characteristics, and learning management system activity (Almalawi et al., 2024; Liz-Domínguez et al., 2019).

Recent developments in explainable artificial intelligence (XAI) have sought to improve transparency and support educator trust in predictive systems (Feldman-Maggor et al., 2025; Saqr & López-Pernas, 2024). Methods such as SHapley Additive



Except where otherwise noted, content in this journal is licensed under a [Creative Commons Attribution 4.0 International Licence](https://creativecommons.org/licenses/by/4.0/). As an open access journal, articles are free to use with proper attribution. ISSN: 2205-0795

exPlanations (SHAP) enable educators to understand the factors contributing to individual predictions and improve the interpretability of learning analytics outputs (Ahmed et al., 2025; Lundberg & Lee, 2017). However, there remains a need for approaches that translate predictive insights into practical tools that support educational decision-making and intervention.

This study was conducted in MAT1114 Introductory Statistics, a first-year introductory statistics unit offered at Edith Cowan University. Using assessment and enrolment data collected during the teaching period, predictive models were developed to identify students at risk of failing the unit. To support practical implementation, model outputs were integrated into a colour-coded risk reporting framework designed to assist educators in prioritising and targeting academic support. The study focused not only on predictive performance but also on the operationalisation of learning analytics within routine educational practice.

Two research questions guided the study:

RQ1: How effectively can pre-examination assessment and enrolment data identify students at risk of failure in a first-year statistics unit?

RQ2: How can explainable learning analytics support educators in understanding and identifying student risk?

Literature Review

Early Identification and Student Success

Research suggests that effective early intervention requires both the timely identification of students experiencing academic difficulty and the provision of actionable information that enables educators to respond appropriately (Liz-Domínguez et al., 2019; Tepgeç & Ifenthaler, 2022). Consequently, learning analytics has emerged as a valuable mechanism for supporting evidence-informed approaches to student success. Learning analytics-informed interventions have the potential to support student achievement, retention, and progression when predictive information is translated into meaningful educational action (Wise, 2014).

Learning Analytics and Early Warning Systems

Learning analytics involves the collection, analysis, and interpretation of educational data for the purpose of understanding and improving learning processes (Brooks & Thompson, 2022). Within higher education, learning analytics has been widely applied to develop early warning systems that identify students at risk of poor academic outcomes using assessment, enrolment, demographic, and behavioural data (Almalawi et al., 2024; Namoun & Alshanqiti, 2021). The value of such systems lies in their ability to identify students early enough for meaningful intervention while there remains sufficient time to influence academic outcomes (Riestra-González et al., 2021).

A consistent finding across the literature is that early assessment performance is among the strongest predictors of subsequent academic success. Formative assessments provide insight into students' developing conceptual understanding and often reveal learning difficulties well before final examinations occur (Chen & Zhai, 2023; Yağcı, 2022). In first-year university units, where extensive prior academic information may not be available, early assessment results provide an important opportunity for identifying students who may benefit from additional support (Goren et al., 2024; Liz-Domínguez et al., 2019).

Predictive Analytics in Higher Education

Machine learning has become a widely used approach for predicting student performance within learning analytics research. Studies have demonstrated that predictive models can classify student outcomes with reasonable accuracy across a variety of educational settings using combinations of academic, demographic, and engagement variables (Butt et al., 2023; Khairy et al., 2024). Tree-based ensemble approaches, including Random Forest, XGBoost, and LightGBM, have frequently demonstrated strong predictive performance because they are capable of capturing complex relationships within educational datasets (Khairy et al., 2024; Wang & Luo, 2024).

Despite these advances, many predictive analytics studies have remained focused on technical model performance and algorithm comparison. Relatively little attention has been given to how predictive outputs are translated into practical tools

that support educational decision-making and intervention (Ramaswami et al., 2023; Susnjak, 2023). As a result, there remains a need for studies that demonstrate how predictive analytics can be operationalised within routine teaching practice.

Explainability and Educator Decision-Making

As predictive systems become increasingly embedded within educational environments, explainability has emerged as an important consideration for responsible implementation. Educators may be reluctant to act on predictions generated by systems they do not understand, particularly when those predictions influence decisions regarding student support and intervention (Feldman-Maggor et al., 2025; Saqr & López-Pernas, 2024).

XAI techniques aim to increase transparency by identifying the factors contributing to model predictions (Ahmed et al., 2025; Guevara-Reyes et al., 2025). SHAP has become one of the most widely adopted approaches because it provides theoretically grounded and interpretable explanations at both the global and individual levels (Lundberg & Lee, 2017; Zhao et al., 2026). At the global level, SHAP identifies the variables that contribute most strongly to predictions across the entire cohort, whereas individual-level explanations show how specific variables influence a particular student's prediction. Within educational contexts, explainability can enhance educator trust, support more informed intervention decisions, and reduce concerns associated with black-box predictive systems (Ahmed et al., 2025; Guevara-Reyes et al., 2025).

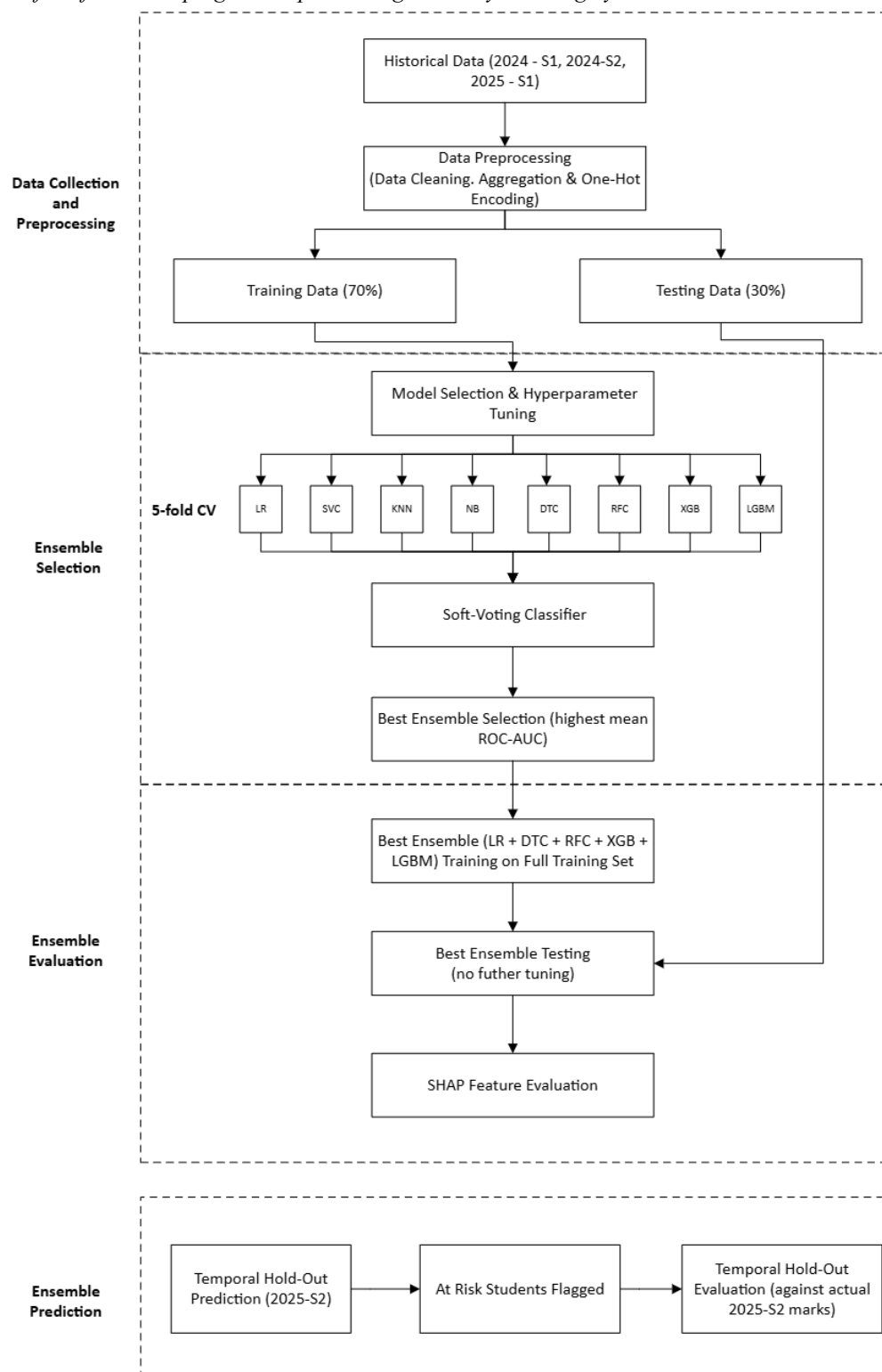
Gaps in the Literature

While substantial research has examined predictive modelling of student performance, two important gaps remain. First, many studies have prioritised predictive accuracy over educational implementation, offering limited guidance on how predictive outputs should inform intervention and student support practices (Ramaswami et al., 2023; Tepgeç & Ifenthaler, 2022). Second, although explainability is increasingly recognised as important, there are relatively few examples of educator-facing systems that integrate predictive modelling and explainable analytics within routine teaching workflows (Feldman-Maggor et al., 2025; Saqr & López-Pernas, 2024).

This study addresses these gaps through the development and evaluation of an interpretable early warning system for a first-year statistics unit. By combining predictive analytics, SHAP-based explanations, and a practical colour-coded risk reporting framework, the study demonstrates how learning analytics can be translated into actionable tools that support educator decision-making and student success.

Methods

Figure 1 provides an overview of the study workflow. The process was implemented in Python and consisted of data collection and preparation, predictive model development, model evaluation, temporal validation, and the generation of an educator-facing risk report to support targeted intervention before the final examination period.

Figure 1*Workflow for Developing and Implementing the Early Warning System*

Note. Student data were collected and prepared, predictive models were developed and evaluated, and predictions were translated into an educator-facing risk reporting framework to support targeted intervention before the final examination period.

Context and Participants

This study was conducted in MAT1114 Introductory Statistics, a first-year introductory statistics unit offered at Edith Cowan University. The unit is a common foundation unit across several undergraduate programs and enrolls students from diverse disciplines, including science, health, education, computing, engineering, and aviation. Early identification of students experiencing academic difficulty is especially important in this setting because large enrolments can reduce opportunities for educators to monitor student progress on an individual basis (Kahu & Nelson, 2018; Kift et al., 2010).

Data used for model development were collected from 566 students enrolled across three consecutive teaching periods: Semester 1 2024 ($n = 214$), Semester 2 2024 ($n = 152$), and Semester 1 2025 ($n = 200$). Students who did not complete one or both formative assessments were excluded because the predictive framework relied on assessment performance as a primary predictor, and meaningful assessment-based predictions could not be generated when these data were unavailable. Although assessment non-submission may itself represent a strong indicator of academic risk, incorporating non-submission as a predictor would require a different modelling approach capable of distinguishing between risk arising from academic performance and risk arising from disengagement. The objective of the present study was to evaluate the predictive value of available assessment and enrolment data among students who participated in the formative assessments. Future iterations of the framework should explicitly incorporate assessment non-submission as a separate risk indicator.

An additional cohort from Semester 2 2025 ($n = 114$ after exclusions) was reserved exclusively for temporal validation. The cohort was not used during model development, model selection, or internal evaluation and was used to assess the framework's performance under authentic deployment conditions.

Predictor Variables

The early warning system was developed using information routinely available before the final examination period. Predictor variables included Test 1 and Test 2 results together with enrolment characteristics comprising course code, study mode, attendance type, and fee category. Course code identified the student's degree program; study mode indicated whether the student was enrolled on a full-time or part-time basis; attendance type identified whether the student studied on campus, online, or at the university's Sri Lanka campus; and fee category distinguished domestic, international fee-paying, and exchange students. Formative assessment results were included because previous research has consistently identified assessment performance as one of the strongest predictors of academic success (Chen & Zhai, 2023; Yağcı, 2022). Enrolment characteristics were included because demographic and enrolment-related variables have frequently been incorporated into learning analytics models as contextual predictors of student outcomes (Almalawi et al., 2024; Namoun & Alshanqiti, 2021).

The outcome variable was final unit status, classified as Pass or Fail according to university grading requirements, where a final mark of 50% or greater was classified as Pass and a final mark below 50% was classified as Fail.

Predictive Modelling

A learning analytics framework was developed to estimate students' likelihood of successfully completing the unit. Eight supervised machine learning algorithms were evaluated: Logistic Regression (LR), Support Vector Classifier (SVC), K-Nearest Neighbours (KNN), Gaussian Naïve Bayes (NB), Decision Tree Classifier (DTC), Random Forest Classifier (RFC), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). These algorithms were selected because they represent a diverse range of modelling approaches commonly used in educational prediction research (Butt et al., 2023; Khairy et al., 2024).

To maximise reproducibility, preprocessing pipelines and hyperparameter settings were fixed prior to model evaluation and applied consistently across all ensemble configurations. Table 1 summarises the machine-learning algorithms, preprocessing pipelines, and model settings used in the study.

All possible non-empty soft-voting ensemble combinations of the eight algorithms were evaluated, resulting in 255 ($2^8 - 1$) candidate ensemble models. Soft-voting ensembles were selected because they combine predicted probabilities from multiple classifiers and leverage the complementary strengths of diverse modelling approaches (Butt et al., 2023; Khairy et al., 2024).

Table 1*Pipelines and Key Hyperparameters (Exact Settings)*

Model	Type	Pipeline	Key Settings
Logistic Regression (LR)	Linear classifier	StandardScaler → LogisticRegression	max_iter=1000, class_weight='balanced'
Support Vector Classifier (SVC)	Kernel-based classifier	StandardScaler → SVC	gamma='auto', C=1, probability=True, class_weight='balanced'
K-Nearest Neighbours (KNN)	Instance-based learner	StandardScaler → KNeighborsClassifier	n_neighbors=5
Naïve Bayes (NB)	Probabilistic classifier	StandardScaler → GaussianNB	None
Decision Tree Classifier (DTC)	Tree-based model	None	class_weight='balanced', random_state=42
Random Forest Classifier (RFC)	Ensemble of decision trees	None	n_estimators=100, class_weight='balanced', random_state=42
XGBoost Classifier (XGB)	Gradient boosting ensemble	None	eval_metric='logloss', random_state=42, scale_pos_weight=scale_pos, max_depth=3, n_estimators=50
LightGBM Classifier (LGBM)	Gradient boosting ensemble	None	class_weight='balanced', random_state=42, min_child_samples=5, max_depth=3, n_estimators=50

Note. *scale_pos_weight* was set using the class ratio (negative_class_count/positive_class_count) in training folds (symbolised as *scale_pos* throughout this study). Unless specified, *learning_rate* followed library defaults; seeds fixed at 42.

Model development used a stratified 70:30 train-test split to preserve the distribution of Pass and Fail outcomes. Candidate ensembles were evaluated using stratified five-fold cross-validation (Kohavi, 1995), with Receiver Operating Characteristic Area Under the Curve (ROC-AUC) used as the primary metric for model selection. The highest-performing ensemble, comprising LR, DTC, RFC, XGBoost, and LightGBM, was subsequently retrained on the full training dataset and evaluated on the independent hold-out dataset.

Although the highest-performing ensemble was selected from 255 candidate configurations using five-fold cross-validation, selecting a single best-performing model introduces some risk of optimistic performance estimation. To mitigate this risk, the selected ensemble was subsequently evaluated on both an independent hold-out dataset and a temporally independent cohort that was not used during model development, model selection, or internal evaluation. Performance was relatively stable among the highest-ranked ensemble configurations. The top five ensembles achieved ROC-AUC values ranging from 0.9146 to 0.9179 and accuracy values ranging from 0.8357 to 0.8484, indicating that predictive performance was not dependent on a single model configuration but was consistently achieved across several ensemble combinations. This suggests that the selected ensemble represents one of several comparably strong models rather than an isolated high-performing solution.

Model Evaluation

Predictive performance was evaluated using ROC-AUC as the primary performance metric because it measures a model's ability to distinguish between students who passed and failed independent of a specific classification threshold and is relatively robust to class imbalance (Almalawi et al., 2024; Namoun & Alshantiti, 2021). Accuracy, precision, recall, and F1-score were also examined to provide a broader assessment of classification performance and to evaluate the balance between correctly identifying students at risk and minimising false classifications (Almalawi et al., 2024; Namoun & Alshantiti, 2021).

Temporal Validation

To assess the generalisability of the predictive framework, the selected ensemble was applied to an independent Semester 2 2025 cohort ($n = 114$) that was not used during model development, model selection, or internal testing. This temporal validation provided an assessment of performance on a future cohort under authentic deployment conditions and allowed evaluation of the model's practical utility for routine teaching and student support activities. Nine students were excluded from prediction because they did not complete one or both formative assessments, which formed the basis of the predictive framework. All nine subsequently did not complete the final examination, and two were subject to academic misconduct investigations. Although insufficient assessment data prevented these students from being included in the prediction process, future implementations of the approach could incorporate referral pathways to university student support services for students exhibiting similar patterns of non-submission. This finding suggests that assessment non-submission may itself represent an important indicator of academic risk and should be considered in future iterations of the framework.

Explainability and Risk Reporting

To improve transparency and support educator trust, SHAP was used to examine the contribution of predictor variables to model outputs (Lundberg & Lee, 2017). SHAP values provide theoretically grounded explanations of model predictions and enable interpretation at both cohort and individual student levels (Ahmed et al., 2025; Lundberg & Lee, 2017). At the cohort level, SHAP can be used to identify the variables that contribute most strongly to prediction outcomes across all students, whereas individual-level explanations show how specific variables contribute to a particular student's risk classification. For example, a student may be classified as High Risk primarily because of low Test 1 and Test 2 results, while another student may receive a similar classification based on a combination of assessment performance and contextual enrolment factors. This approach helped educators understand why students were identified as being at risk and which variables contributed most strongly to individual predictions.

Prediction outputs were translated into an educator-facing colour-coded risk reporting framework. Students were grouped according to predicted probability of failure, enabling teaching staff to identify students requiring support and prioritise intervention efforts.

Operational Implementation and Educator Use

To support practical implementation, an operational workflow was developed to translate model predictions into actionable information for educators.

The framework was designed to operate using information routinely available within university information systems. Test 1 and Test 2 assessment results were extracted from the Canvas learning management system, while enrolment characteristics, including course code, study mode, attendance type, and fee category, were obtained from the university student information system. Test 1 was administered in Week 6 and Test 2 in Week 10 of the teaching period. These variables were selected because they were available before the final examination period, enabling identification of students who may benefit from additional support. Predictions generated following Test 2 typically provided approximately five weeks before the final examination period for intervention.

Data were exported from Canvas and the student information system as comma-separated value (CSV) files and converted into Microsoft Excel format for processing. Student records were matched using institutional student identification numbers, and the required predictor variables were consolidated into a single dataset. This dataset was then processed using the workflow shown in Figure 1 to generate predicted probabilities of failure and corresponding risk classifications.

Following Test 2, the trained ensemble model generated an educator-facing Excel risk report containing student identifiers, predicted probabilities of failure, risk classifications, and explanatory information derived from SHAP analysis. This enabled teaching staff to identify students who may require additional support before final assessment activities.

Although the implementation described in this study utilised Canvas and the university student information system, the framework was not dependent on a specific platform. The required inputs consisted primarily of assessment results and enrolment information that are commonly available within learning management and student administration systems.

Consequently, the framework could be adapted to other environments, including Moodle, Blackboard, and comparable institutional systems, provided equivalent data were available.

The Excel risk report was intended to support educator decision-making rather than replace professional judgement. Teaching staff could use risk classifications to prioritise outreach and allocate available support resources. For example, students classified as Priority Risk could be contacted individually or referred to academic support services, while students classified as High Risk could receive targeted academic support resources and ongoing monitoring. In this way, predictive outputs were translated into a practical workflow that could be incorporated into routine teaching and student support activities.

Results

Predictive Performance

The predictive modelling framework demonstrated strong performance in identifying students at risk of failure. Performance metrics for both the internal hold-out dataset and the temporally independent cohort are presented in Table 2.

Table 2

Summary of Predictive Performance for the Selected Ensemble Model Across the Internal Hold-Out and Temporally Independent Cohorts

Metric	Internal Hold-Out	Temporal Cohort
ROC-AUC	0.918	0.851
Accuracy	0.836	0.737
Precision	0.780	0.561
Recall	0.797	0.865
F1-score	0.784	0.681

Across all ensemble configurations evaluated, predictive performance was consistently high, indicating that formative assessment and enrolment data provided valuable indicators of academic outcomes. As shown in Table 2, the selected ensemble achieved a ROC-AUC of 0.918, an accuracy of 0.836, a precision of 0.780, a recall of 0.797, and an F1-score of 0.784 on the internal hold-out dataset. When evaluated on a temporally independent cohort not used during model development, the model achieved a ROC-AUC of 0.851, an accuracy of 0.737, a precision of 0.561, a recall of 0.865, and an F1-score of 0.681, demonstrating strong generalisability under authentic deployment conditions.

The relatively high recall observed in the temporal cohort indicates that the framework successfully identified most students who subsequently experienced academic difficulty. Of the 37 students who ultimately failed the unit, 32 were correctly identified by the model, while five students were not identified and therefore represent false negatives. Although some students were incorrectly classified as at risk, the framework captured the majority of future fail outcomes, which is particularly important in an early warning context where missed at-risk students may have limited opportunity to receive support. From an intervention perspective, the relatively high recall was considered desirable because failing to identify students requiring support may have greater educational consequences than providing additional support to some students who subsequently pass.

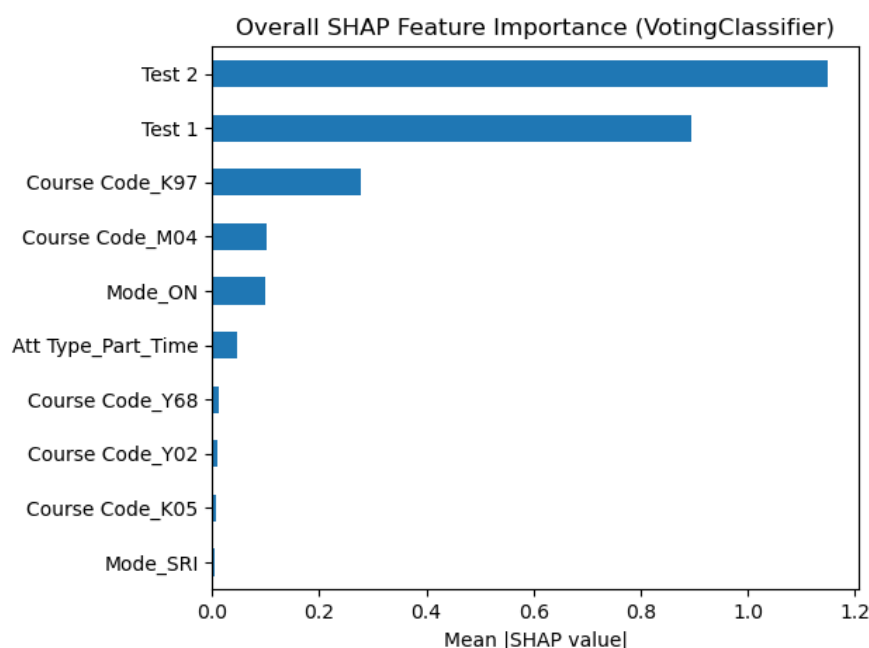
These findings indicate that routinely collected assessment and enrolment data can support the identification of students who may benefit from academic support before final assessment activities are undertaken.

Factors Associated with Student Success

To improve interpretability, SHAP analysis was used to identify the variables contributing most strongly to student outcomes. The relative importance of predictor variables is presented in Figure 2.

Figure 2

Mean Absolute SHAP Values Showing the Relative Contribution of Predictor Variables to Student Outcome Predictions



Note. Test 2 and Test 1 were the strongest predictors of student success, substantially outweighing enrolment-related characteristics.

As shown in Figure 2, Test 2 and Test 1 were the strongest contributors to prediction outcomes, with substantially higher mean SHAP values than all other variables. Course code and study mode contributed additional contextual information, but their influence was considerably smaller. These findings indicate that formative assessment performance provides the clearest signal of subsequent academic success, with enrolment-related characteristics making comparatively modest contributions to prediction outcomes.

The findings suggest that students experiencing difficulty in the formative assessments are significantly more likely to experience academic challenges later in the semester. Consistent with previous research, formative assessment performance provided the strongest indication of subsequent academic success and therefore represents a valuable source of information for identifying students who may benefit from additional support (Chen & Zhai, 2023; Yağcı, 2022).

Consistent with previous learning analytics research, formative assessment performance emerged as the strongest predictor of final unit outcomes. The contribution of the present study is therefore not the identification of this relationship itself, as the predictive value of formative assessment performance has been well established in the literature (Chen & Zhai, 2023; Goren et al., 2024; Yağcı, 2022). Rather, the study demonstrates how these predictive signals can be translated into an interpretable and operational framework that supports educator decision-making. By combining temporal validation, explainable analytics, and an educator-facing risk reporting workflow, the framework illustrates how learning analytics can be integrated into routine student support practices within large first-year cohorts.

Early Warning and Intervention Framework

Prediction outputs were translated into a colour-coded risk reporting framework that enabled teaching staff to identify students requiring support. The intervention categories used by the framework are summarised in Table 3. Risk categories were defined

using the model's predicted final unit outcome (Pass or Fail) together with predicted probabilities of failure and were linked to corresponding intervention priorities.

Table 3

Risk Categories Used within the Early Warning Framework

Risk Level	Criteria	Suggested Response
Priority Risk (Dark Red)	Highest-risk 20% of students with predicted Fail outcome and probability ≥ 0.70	Individual outreach, referral to academic support services, consultation appointment
High Risk (Red)	Remaining students with predicted Fail outcome and probability ≥ 0.70	Targeted academic support resources, ongoing monitoring
Low Risk (Green)	Predicted Fail outcome and probability < 0.70 , or predicted Pass outcome	No intervention required

Note. The 0.70 intervention threshold was selected to support targeted allocation of limited support resources and should be interpreted as an operational rather than a universal threshold.

To support implementation, prediction outputs were translated into an educator-facing Excel risk report. Figure 3 presents an example of the reporting interface used by teaching staff.

Figure 3

Example of the Educator-Facing Excel Risk Report

Unit Class	Course Code	Att Type	Fee Category	Test 1	Test 2	Predicted Grade	Predicted Probability	At-Risk
ON	M04	Full_Time	UGCS	13.33	17.59	Pass	0.63	No
ON	K97	Full_Time	UGCS	20.83	20.37	Pass	0.82	No
ON	X05	Full_Time	EXCHANGE	22.50	20.51	Pass	0.99	No
OFF	K97	Part_Time	UGCS	8.33	2.78	Fail	0.79	Yes
OFF	M04	Full_Time	UGCS	12.08	17.59	Fail	0.62	No
OFF	Y02	Full_Time	UGCS	14.17	13.89	Fail	0.84	Yes
OFF	U65	Full_Time	UGCS	15.83	12.96	Pass	0.52	No
ON	K05	Full_Time	UGCS	7.50	2.78	Fail	0.97	Yes
OFF	K97	Full_Time	UGCS	18.33	12.04	Fail	0.58	No

Note. Students are automatically classified into risk categories using predicted probabilities, enabling rapid identification and prioritisation of students requiring academic support.

Students predicted to fail with a predicted probability of 0.70 or greater were identified as requiring intervention. To assist educators in prioritising outreach efforts, these students were ranked according to their predicted probability. The highest-risk 20% of students within this intervention group were classified as Priority Risk and highlighted in dark red. The remaining students predicted to fail with a probability of 0.70 or greater were classified as High Risk and highlighted in red. Students predicted to fail with probabilities below 0.70, together with all students predicted to pass, were classified as Low Risk and highlighted in green and were not targeted for intervention.

A threshold of 0.70 was selected pragmatically to focus intervention resources on students with the highest estimated likelihood of academic failure while maintaining a manageable number of students requiring outreach and support. The purpose of the framework was not to intervene with every student identified as potentially at risk, but rather to prioritise students most likely to require immediate assistance. By concentrating support efforts on students with the highest predicted

probability of failure, the framework enabled teaching staff to allocate limited intervention resources efficiently while maintaining a practical workload for follow-up activities. This colour-coded approach enables educators to rapidly identify and prioritise students most likely to experience academic difficulty, supporting more efficient allocation of intervention resources.

Application of the framework to the temporal validation cohort ($n = 114$) demonstrated its practical value. Of the 37 students classified as Priority Risk or High Risk, 23 (62.2%) subsequently failed the unit, indicating that the framework captured a substantial proportion of future fail outcomes within a relatively small intervention group. The remaining 14 students, all classified as High Risk, subsequently passed the unit, indicating that risk classifications should be interpreted as indicators of elevated academic risk rather than deterministic predictions of failure.

This finding suggests that the framework can assist educators in directing intervention efforts towards a relatively small group of students while concentrating support resources on those most likely to experience academic difficulty.

The resulting risk report provided staff with a concise and interpretable overview of student risk, enabling intervention decisions to be incorporated into routine teaching and student support processes.

Discussion

The findings demonstrate that routinely collected assessment and enrolment data can be used to identify students at risk of academic difficulty before the final examination period. The selected ensemble achieved strong predictive performance on both the internal hold-out dataset (ROC-AUC = 0.918) and the temporally independent validation cohort (ROC-AUC = 0.851), indicating that model performance remained robust when applied to a future student cohort. These findings suggest that early warning systems can provide educators with a practical mechanism for identifying students who may benefit from additional academic support before the final assessment. Such early identification is particularly important in first-year units, where academic difficulties often emerge during the transition to university study and may contribute to disengagement, poor performance, and attrition.

A key finding of the study is the importance of formative assessment performance as an indicator of academic success. SHAP analysis revealed that Test 1 and Test 2 were substantially stronger predictors of final outcomes than enrolment-related characteristics, suggesting that formative assessment performance during the teaching period provides a reliable indication of subsequent academic success. Consistent with previous learning analytics research, formative assessment performance emerged as the strongest predictor of final unit outcomes; therefore, the contribution of the present study is not the identification of this relationship itself, as its predictive value has been well established in the literature (Chen & Zhai, 2023; Goren et al., 2024; Yağcı, 2022). Rather, the study demonstrates how these predictive signals can be translated into an interpretable and operational framework that supports educator decision-making through temporal validation, explainable analytics, and an educator-facing risk-reporting workflow.

Beyond predictive accuracy, the study demonstrates how learning analytics can be translated into educational practice. The educator-facing risk report enabled teaching staff to identify students requiring support following completion of Test 2 and to prioritise intervention efforts through the Priority Risk and High Risk classifications. Importantly, predictive models alone do not improve student outcomes; educational benefit is realised only when predictive insights are integrated into teaching and student support processes. The framework therefore illustrates how learning analytics can be operationalised within routine educational workflows.

The inclusion of explainable analytics provides insight into how predictive outputs can be translated into information that educators can understand and act upon. Rather than relying on opaque predictions, educators are provided with information about the factors influencing a student's risk classification. This transparency may support greater confidence in the use of learning analytics and facilitate more informed conversations between educators and students regarding academic progress and support needs. Explainability is particularly important in educational settings, where decisions affecting students should be transparent, justifiable, and ethically responsible. In this study, SHAP explanations highlighted the dominant influence of formative assessment performance on risk classification, enabling educators to distinguish between classifications driven primarily by assessment outcomes and those influenced by contextual enrolment characteristics. Such insights may support more informed decision-making and help educators tailor intervention strategies to individual student needs. For example,

educators could distinguish between students identified primarily because of low assessment performance and those whose classifications were influenced by a combination of assessment and contextual factors, providing additional context for support discussions.

Although enrolment characteristics contributed to the predictive framework, SHAP analysis indicated that their influence was substantially smaller than that of formative assessment performance. These contextual variables should not be interpreted as indicators of student capability nor as causal explanations of academic risk. Rather, they provide supplementary contextual information that may assist interpretation alongside observed academic behaviours. Importantly, the framework was designed to support targeted outreach and resource allocation rather than high-stakes decisions regarding progression or grading. Educator judgement therefore remains essential to minimise potential bias and avoid unintended stigmatisation of particular student groups.

The findings also have implications for institutional student success and retention strategies. In large first-year units, academic support resources are often limited, requiring educators to prioritise intervention efforts. The 0.70 probability threshold was selected to concentrate intervention resources on students with the highest estimated likelihood of academic failure while maintaining a manageable intervention workload. However, students at moderate levels of risk may also benefit from support, and alternative thresholds may be appropriate depending on cohort size, intervention objectives, and available resources. Lower thresholds would increase the number of students identified for intervention and improve coverage of at-risk students, whereas higher thresholds would reduce intervention workload while focusing resources on students with the highest estimated likelihood of academic difficulty. Consequently, intervention thresholds should be viewed as operational decisions rather than universal recommendations.

In the temporal validation cohort, 23 of 37 students (62.2%) classified as Priority Risk or High Risk subsequently failed the unit, suggesting that the framework successfully concentrated a substantial proportion of future fail outcomes within a relatively small intervention group. The remaining 14 students, all classified as High Risk, subsequently passed the unit, highlighting that prediction-based classifications should be viewed as indicators of elevated risk rather than definitive academic outcomes. Together, these findings suggest that the framework can assist educators in directing intervention efforts towards students most likely to experience academic difficulty while retaining flexibility to adapt intervention strategies to local institutional contexts.

Several limitations should be acknowledged. First, the study was conducted within a single first-year statistics unit at one Australian university, which may limit the generalisability of the findings to other disciplines or institutions. Second, the predictive framework relied primarily on assessment and enrolment data. While these variables proved effective, the inclusion of additional indicators such as learning management system engagement, attendance patterns, or assessment submission behaviour may further enhance predictive capability. Notably, students who did not complete one or both formative assessments could not be included in the prediction framework because insufficient assessment data were available. In the temporal validation cohort, all students excluded for this reason subsequently failed to complete the final examination. This suggests that assessment non-submission may itself represent an important indicator of academic risk and should be incorporated into future early warning systems. Third, the study evaluated the ability to identify students at risk rather than the effectiveness of subsequent interventions. Furthermore, not all students classified as High Risk subsequently failed, highlighting that risk classifications identify students with elevated likelihoods of academic difficulty rather than certainty of failure. Further research is therefore required to determine which students are most likely to benefit from intervention and which intervention strategies are most effective for different levels of student risk.

Future research should investigate how educators use risk information in practice and examine whether targeted interventions informed by early warning systems lead to measurable improvements in student outcomes, retention, and progression. Additional research is also needed to determine which students are most likely to benefit from intervention and whether alternative intervention thresholds may provide a more effective balance between support effort and educational impact. Further research could also explore the incorporation of assessment non-submission behaviour as an explicit predictor within future early warning frameworks.

Conclusion

This study developed and evaluated an interpretable early warning system for supporting student success in a first-year statistics unit. Using routinely collected assessment and enrolment data, the framework successfully identified students at risk of academic difficulty before the final examination period and maintained strong predictive performance when evaluated on an independent temporal validation cohort.

A key contribution of the study is the translation of predictive analytics into an educator-facing intervention framework. By combining risk prediction, explainable analytics, and an Excel-based reporting tool, the system provides educators with actionable information that can be used to identify students requiring support, prioritise intervention efforts, and inform decision-making. The integration of SHAP explanations further enhances transparency by helping educators understand the factors contributing to individual risk classifications.

The findings demonstrate that learning analytics can play a valuable role in supporting evidence-informed student success initiatives when predictive insights are presented in ways that are transparent, interpretable, and practical for educators. More broadly, the study illustrates how predictive analytics can be operationalised within routine teaching and student support workflows to facilitate timely intervention in large first-year cohorts while maintaining transparency and interpretability for educators.

Disclosure Statements

1. *Ethics Approval and Data Privacy*

This study received approval from the Edith Cowan University Human Research Ethics Committee (HREC Approval ID: 2025-07193-Lam). All student data were anonymised prior to analysis and handled in accordance with university data governance and privacy requirements. Predictions generated by the model were used solely to support academic intervention and student support activities and were not used for grading, progression, or other high-stakes decisions.

2. *Data Availability*

The dataset analysed in this study contains confidential student information and cannot be made publicly available under institutional and ethical requirements. However, the Python workflow used to generate the predictive analytics and educator-facing risk reports is available at: https://github.com/blam-ecu/student_success.git. Educators may adapt the workflow using their own institutional data.

De-identified metadata or derived materials may be made available by the corresponding author upon reasonable request, subject to ethics approval and institutional data governance requirements.

3. *Conflict of Interest*

The authors declare no conflicts of interest in relation to the research, authorship, and publication of this article.

4. *Funding*

The authors received no financial support for the research, authorship, and/or publication of this article.

5. *Use of Artificial Intelligence Tools*

AI-assisted tools (Microsoft Copilot) were used solely to assist with language editing, proofreading, and manuscript refinement. All study design, data analysis, and results interpretation were conducted by the authors.

6. *Author Contributions*

BL: Conceptualisation, methodology, formal analysis, investigation, writing — original draft preparation, writing — review and editing.

SR: Writing — review and editing.

References

- Ahmed, W., Wani, M. A., Plawiak, P., Meshoul, S., Mahmoud, A., & Hammad, M. (2025). Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions. *Scientific Reports*, 15(1), Article 26879. <https://doi.org/10.1038/s41598-025-12353-4>
- Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive models for educational purposes: A systematic review. *Big Data and Cognitive Computing*, 8(12), Article 187. <https://doi.org/10.3390/bdcc8120187>
- Brooks, C., & Thompson, C. (2022). Predictive modelling in teaching and learning. In C. Lang, G. Siemens, A. Wise, D. Gašević, & A. Merceron (Eds.), *Handbook of learning analytics* (2nd ed., pp. 31–42). Society for Learning Analytics Research. <https://doi.org/10.18608/hla22.003>
- Butt, N. A., Mahmood, Z., Shakeel, K., Alfarhood, S., Safran, M., & Ashraf, I. (2023). Performance prediction of students in higher education using multi-model ensemble approach. *IEEE Access*, 11, 136091–136108. <https://doi.org/10.1109/ACCESS.2023.3336987>
- Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, 28(9), 12039–12057. <https://doi.org/10.1007/s10639-023-11672-1>
- Feldman-Maggor, Y., Cukurova, M., Kent, C., & Alexandron, G. (2025). The impact of explainable AI on teachers' trust and acceptance of AI EdTech recommendations: The power of domain-specific explanations. *International Journal of Artificial Intelligence in Education*, 35(5), 2889–2922. <https://doi.org/10.1007/s40593-025-00486-6>
- Goren, O., Cohen, L., & Rubinstein, A. (2024). Early prediction of student dropout in higher education using machine learning models. In *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 349–359). International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.12729834>
- Guevara-Reyes, R., Ortiz-Garcés, I., Andrade, R., Cox-Riquetti, F., & Villegas-Ch, W. (2025). Machine learning models for academic performance prediction: Interpretability and application in educational decision-making. *Frontiers in Education*, 10, Article 1632315. <https://doi.org/10.3389/feduc.2025.1632315>
- Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: Understanding the mechanisms of student success. *Higher Education Research & Development*, 37(1), 58–71. <https://doi.org/10.1080/07294360.2017.1344197>
- Khairy, D., Alharbi, N., Amasha, M. A., Areed, M. F., Alkhalaf, S., & Abougalala, R. A. (2024). Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*, 29, 21621–21645. <https://doi.org/10.1007/s10639-024-12619-w>
- Kift, S., Nelson, K., & Clarke, J. (2010). Transition pedagogy: A third generation approach to FYE: A case study of policy and practice for the higher education sector. *The International Journal of the First Year in Higher Education*, 1(1), 1–20. <https://doi.org/10.5204/intjfyhe.v1i1.13>
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137–1143). Morgan Kaufmann. <https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf>
- Liz-Domínguez, M., Caeiro-Rodríguez, M., Llamas-Nistal, M., & Mikic-Fonte, F. (2019). Predictors and early warning systems in higher education: A systematic literature review. *CEUR Workshop Proceedings*, 2415, 84–99. <https://www.ceur-ws.org/Vol-2415/paper08.pdf>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. https://papers.nips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), Article 237. <https://doi.org/10.3390/app11010237>
- Ramaswami, G., Susnjak, T., & Mathrani, A. (2023). Effectiveness of a learning analytics dashboard for increasing student engagement levels. *Journal of Learning Analytics*, 10(3), 115–134. <https://doi.org/10.18608/jla.2023.7935>
- Riestra-González, M., del Puerto Paule-Ruiz, M., & Ortín, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, 163, Article 104108. <https://doi.org/10.1016/j.compedu.2020.104108>
- Sagr, M., & López-Pernas, S. (2024). Why explainable AI may not be enough: Predictions and mispredictions in decision making in education. *Smart Learning Environments*, 11, Article 52. <https://doi.org/10.1186/s40561-024-00343-4>
- Susnjak, T. (2023). *A prescriptive learning analytics framework: Beyond predictive modelling and onto explainable AI with prescriptive analytics and ChatGPT*. arXiv. <https://arxiv.org/html/2208.14582>
- Tepgeç, M., & Ifenthaler, D. (2022). Learning analytics based interventions: A systematic review of experimental studies. In D. G. Sampson, D. Ifenthaler, & P. Isaias (Eds.), *Proceedings of the International Conference on Cognition and Exploratory Learning in the Digital Age 2022* (pp. 327–330). International Association for Development of the Information Society. <https://files.eric.ed.gov/fulltext/ED626950.pdf>

- Wang, S., & Luo, B. (2024). Academic achievement prediction in higher education through interpretable modeling. *PLoS ONE*, 19(9), Article e0309838. <https://doi.org/10.1371/journal.pone.0309838>
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In *LAK '14: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 203–211). Association for Computing Machinery. <https://doi.org/10.1145/2567574.2567588>
- Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9, Article 11. <https://doi.org/10.1186/s40561-022-00192-z>
- Zhao, C., Liu, J., & Parilina, E. (2026). The Shapley value contribution to explainable artificial intelligence: A comprehensive survey. *Dynamic Games and Applications*, 16(3), 1013–1050. <https://doi.org/10.1007/s13235-025-00670-2>

Please cite this article as:

Lam, B., & Richardson, S. (2026). Supporting student success through an interpretable early warning system in a first-year statistics unit. *Student Success*. Advance online publication. <https://doi.org/10.63608/ssj.4175>

This article has been accepted for publication in *Student Success*. Please see the Editorial Policies under the 'About' section of the Journal website for further information.

Student Success: A journal exploring the experiences of students in tertiary education.



Except where otherwise noted, content in this journal is licensed under a [Creative Commons Attribution 4.0 International Licence](https://creativecommons.org/licenses/by/4.0/). As an open access journal, articles are free to use with proper attribution. ISSN: 2205-0795