Governed Early-Warning Analytics for Student Success in Digital Higher Education: A Business-Oriented Evidence Model

Nahla Moussa[1,*], **Low Hon Loon Alfred**[2]

[1]Associate Professor, Amity University Dubai, UAE

[2]Senior Manager (Instructional Design), Higher Colleges of Technology, Abu Dhabi, UAE

Emails: nahlamoussa2020@gmail.com; lalfred@hct.ac.ae

Abstract

Student-success analytics has moved from experimental prediction toward an institutional capability for reducing attrition, allocating support resources, and improving digital learning governance. This paper develops a business-oriented early-warning model for education technology environments in which predictive performance, interpretability, intervention priority, and governance are treated as joint design requirements. The study uses a public student-success dataset from a higher education institution and evaluates decisive outcome prediction for dropout and graduation, while preserving a wider discussion of the enrolled group as an unresolved operational state. The proposed model combines a transparent predictive layer, a risk-to-action prioritization layer, and a governance layer that restricts how predictions are translated into student support decisions. The results show that a parsimonious logistic specification can provide competitive performance compared with more complex tree-based models, while producing clearer accountability for academic advising and digital student-success units. The discussion argues that student-success technology should not be judged by accuracy alone, but by whether the analytics pipeline produces timely, explainable, privacy-aware, and operationally usable support signals.

Keywords: Education technology; Student success; Learning analytics; Dropout prediction; Early-warning systems; Higher education; Predictive governance

1. Introduction

Digital higher education has increased the amount of data available about students' academic progression, engagement with learning platforms, administrative status, and support needs. These data streams have made student-success analytics a central topic in education technology because universities are under pressure to improve retention while using advising and student-support resources efficiently. However, prediction alone is not enough. A model that identifies at-risk students but does not explain the reason for the risk, prioritize intervention capacity, or satisfy privacy expectations is unlikely to become a trusted institutional process.

The business problem addressed in this paper is therefore not simply whether dropout can be predicted. The problem is how an institution can convert student data into an early-warning portfolio that is accurate enough for operational use, transparent enough for academic decision-making, and restrained enough to avoid harmful automation. Recent learning analytics research has emphasized the importance of learning constructs, stakeholder adoption, feedback quality, and privacy in student-success systems [3, 5, 15, 17]. This study follows that direction by placing the predictive model inside a governed analytics workflow.

The paper contributes a different structure from purely technical student dropout studies. First, it frames the model as a student-success business capability rather than a stand-alone classifier. Second, it uses a public dataset with real student outcomes to provide reproducible empirical values. Third, it introduces a risk-to-action layer that ranks signals by their expected usefulness for academic support, not only by statistical weight. Fourth, it discusses governance issues that are often left outside model performance tables.

The study uses the public "Predict Students' Dropout and Academic Success" dataset, released through the UCI Machine Learning Repository and described in a data article by Realinho and colleagues [1, 2]. The dataset is suitable because it contains enrolment, demographic, socioeconomic, macroeconomic, and curricular performance indicators linked to final student status. It also has no missing values and is explicitly designed for dropout and academic-success classification.

The analysis is organized around a managerial question: how can a university transform predictive evidence into responsible intervention capacity? The proposed answer is a governed early-warning model that combines statistical classification with intervention ranking, actionability weighting, and human review. This structure is consistent with the business literature style used in recent education technology studies, where model value is evaluated through performance, operational relevance, and adoption constraints rather than through predictive accuracy alone.

2. Related Work

Learning analytics has become a mature research stream in education technology, but recent reviews show that its value depends on more than data availability. Ifenthaler and Yau [3] reviewed evidence on the use of learning analytics to support study success in higher education, highlighting continuation and completion as key institutional outcomes. Knobbout and Van Der Stappen [5] showed that learning analytics interventions require careful operationalization of learning-related constructs. Pan et al. [17] further examined learning analytics interventions embedded in learning management systems and concluded that instructional use remains uneven across contexts. These works support a design logic in which prediction is connected to intervention and institutional decision-making.

DOI: https://doi.org/10.54216/IJAIET.040202

A second line of research focuses on early alerts and stakeholder adoption. Atif et al. [4] examined teacher perceptions of early alert systems and showed that usability, trust, and intervention clarity shape adoption. Dashboard-oriented studies similarly emphasize that analytics must be interpretable to educators and students, not merely technically correct [8]. In higher education management, such findings imply that the success of an early-warning system depends on whether academic advisors can understand the risk drivers and act on them within existing student-support processes.

A third stream examines explainability, feedback, and self-regulated learning. Afzaal et al. [6] proposed explainable AI for data-driven feedback and intelligent action recommendations to support students' self-regulation. Karaoglan Yilmaz [9] found that analytics-assisted recommendations and guidance feedback can improve metacognitive awareness and academic achievement, while related studies connected analytics interventions to self-efficacy, problem-solving, and online engagement [10, 11]. Lu et al. [13] added evidence from blended learning, showing that intervention design can influence self-regulated learning outcomes. Together, these studies suggest that a student-success system should offer actionable feedback rather than a static risk score.

A fourth stream focuses on behavioral and process modelling. Tan and Samavedham [12] used LMS sequence analysis to examine procrastination, showing that process timing can matter as much as final grade indicators. Al-Shaikhli et al. [7] linked weekly learning-outcome visualization to continued LMS use and perceived learning self-regulation. These studies are important for education technology because they show that student success is not only a final status but also a process that unfolds through academic behaviors, digital engagement, and institutional support.

Recent empirical studies using public or institutional dropout datasets have evaluated many supervised learning algorithms. Realinho et al. [2] documented the UCI student-success dataset used in this paper. Villar and de Andrade [19] compared supervised algorithms on dropout and academic-success prediction with attention to class imbalance, and Kim et al. [14] emphasized high precision and recall in university dropout prediction. Vaarma and Li [18] used transcript, demographic, and LMS data in Finnish higher education, while Goren et al. [20] evaluated early prediction using machine learning models. Hassan et al. [21] also studied dropout prediction using supervised learning in a national education accessibility survey context. These works provide strong predictive baselines, but their managerial translation often remains underdeveloped.

Governance and privacy form the final foundation of the present study. Francis et al. [15] examined privacy within student-success information systems and showed why institutions must control how student data are interpreted and used. Mukred et al. [16] studied learning analytics tool adoption in higher learning institutions and emphasized decision-making and acceptance factors. These studies indicate that student-success analytics must be evaluated as a socio-technical capability. The model should not only predict risk but also define who receives the alert, what explanation accompanies it, and which governance constraints apply before intervention.

Table 1: Extended positioning of recent studies on education technology and student-success analytics.

Study	Context	Methodological focus	Contribution to the present paper
Ifenthaler and Yau (2020)	Higher education study success	Systematic review	Establishes continuation and completion as central learning analytics outcomes.
Atif et al. (2020)	Early alert systems	Teacher-perspective evaluation	Shows that usefulness, trust, and practical intervention design influence adoption.
Knobbout and Van Der Stappen (2020)	Learning analytics interventions	Construct operationalization review	Motivates transparent measurement rather than black-box risk scoring.
Afzaal et al. (2021)	Data-driven feedback	Explainable AI and action recommendations	Links prediction with feedback that can support student self-regulation.
Al-Shaikhli et al. (2022)	LMS continuation	Weekly learning outcome visualization	Shows how visualization can support learning regulation and continued LMS use.
Jayashanka et al. (2022)	Higher education dashboards	Technology-enhanced dashboard design	Supports the need for interpretable dashboards in institutional settings.
Karaoglan Yilmaz (2022a)	Analytics-assisted feedback	Recommendation and guidance feedback	Demonstrates that analytics-guided feedback can affect metacognition and achievement.
Karaoglan Yilmaz (2022b)	Academic self-efficacy	Learning analytics intervention	Connects analytics use with self-efficacy and problem-solving outcomes.
Karaoglan Yilmaz and Yilmaz (2022)	Online learning engagement	Learning analytics intervention	Supports intervention-oriented design for student engagement.
Tan and Samavedham (2022)	LMS process data	Sequence analysis	Demonstrates the value of temporal process indicators for identifying procrastination.

Table 1: Extended positioning of recent studies on education technology and student-success analytics (continued).

Study	Context	Methodological focus	Contribution to the present paper
Lu et al. (2022)	Blended learning	Quasi-experimental intervention study	Provides evidence that intervention design can influence self-regulated learning.
Realinho et al. (2022)	Public higher education dataset	Dataset documentation and modelling	Provides the empirical dataset and variables used in this study.
Kim et al. (2023)	University dropout prediction	Supervised learning with precision/recall focus	Reinforces the need to evaluate false alerts and missed-risk cases.
Francis et al. (2023)	Student-success information systems	Privacy analysis	Supports the governance layer and limits on automated student-risk use.
Pan et al. (2024)	LMS-based learning analytics	Systematic review of interventions	Shows that analytics impact depends on intervention design and instructional integration.
Mukred et al. (2024)	Higher learning institutions	Learning analytics tool adoption	Shows the relevance of acceptance, decision-making, and institutional readiness.
Vaarma and Li (2024)	Finnish higher education	Machine learning with transcript, demographic, and LMS data	Supports the combination of academic and digital traces for risk prediction.
Villar and de Andrade (2024)	Public dropout and academic success data	Supervised algorithm comparison	Provides a recent benchmark on the same public dataset.
Goren et al. (2024)	Higher education dropout prediction	Early machine learning prediction	Highlights generalizability and timing concerns in early prediction.
Hassan et al. (2024)	National education accessibility survey	Supervised dropout prediction	Extends dropout analytics to broader education access and policy contexts.

3. Proposed Governed Early-Warning Model

The proposed model treats student-success analytics as a governed institutional process. It contains three linked layers: a predictive layer, an intervention-priority layer, and a governance layer. The predictive layer estimates the probability that a student belongs to a successful academic outcome. The intervention-priority layer transforms this estimate into an actionable support score. The governance layer decides whether the alert can be released to advisors, based on explainability and policy constraints.

Let $\mathbf{x}_i \in \mathbb{R}^p$ be the feature vector for student i, where the variables include academic progression, admission route, socioeconomic indicators, tuition status, and curricular performance. The primary model estimates:

$$P(y_i = 1 \mid \mathbf{x}_i) = \sigma(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}), \tag{1}$$

where $y_i = 1$ denotes graduation, $\sigma(z) = 1/(1 + e^{-z})$, β_0 is the intercept, and $\boldsymbol{\beta}$ is the coefficient vector. The dropout risk score is therefore:

$$R_i = 1 - P(y_i = 1 \mid \mathbf{x}_i). \tag{2}$$

The intervention-priority layer converts risk into action. Let A_j denote the estimated actionability of feature j, and let I_j denote its normalized contribution to prediction. The intervention priority for student i is defined as:

$$S_i = R_i \sum_{j=1}^{p} \omega_j A_j I_j |x_{ij} - \mu_j|, \tag{3}$$

where ω_j is a policy weight and μ_j is a reference value for feature j. This term prevents the system from treating all risk factors equally. For example, failed curricular units and unpaid tuition may require different institutional responses.

The governance layer applies a decision rule before any action is taken:

$$D_i = \begin{cases} \text{monitor}, & R_i < \tau_1, \\ \text{advisor review}, & \tau_1 \le R_i < \tau_2, \\ \text{coordinated intervention}, & R_i \ge \tau_2 \text{ and } G_i = 1, \end{cases} \tag{4}$$

where G_i is a governance approval indicator. It equals one only when the data used for the alert are allowed under institutional policy and the explanation is available to the advisor. This design avoids direct automation of high-impact educational decisions.

Algorithm 1 Governed Student Success Early-Warning Model

Require: Student feature matrix X, labels Y, actionability vector A, thresholds τ_1, τ_2
Ensure: Risk band and intervention priority for each student

```
Standardize continuous variables and encode categorical indicators.
Partition X, Y into training and hold-out sets using a stratified split.
Train an interpretable probabilistic classifier to estimate p_i = P(y_i = 1 | x_i).
Compute dropout risk R_i = 1 − p_i for each student in the scoring population.
Estimate feature contribution vector I using standardized coefficients and model inspection.
for each student i do
    Compute intervention score S_i = R_i Σ_j ω_j A_j I_j |x_ij − μ_j|.
    Assign the student to monitor, advisor review, or coordinated intervention.
    Release the alert only if governance approval G_i = 1.
end for
Return ranked intervention list with risk explanation.
```

4. Data and Method

The empirical analysis uses the UCI student-success dataset [1]. It contains 4,424 records and variables describing application mode, course, previous qualification, parental qualification and occupation, tuition status, scholarship status, age at enrolment, curricular units, grades, macroeconomic context, and final status. The three outcome classes are dropout, enrolled, and graduate. Because the enrolled state is not a final success/failure outcome, the main predictive experiment follows a decisive-outcome protocol that removes the enrolled group and compares dropout with graduation. This leaves 3,630 records.

The analysis uses an 80/20 hold-out split consistent with the dataset documentation. Logistic regression is used as the primary governed model because it supports transparent explanation. Random forest and extreme gradient boosting are included as alternative models to test whether additional non-linearity improves operational value. Accuracy, precision, recall, macro F1, weighted F1, and confusion matrices are reported. The paper also presents a tri-class sensitivity benchmark from a recent comparative study on the same public dataset [19].

5. Results

Figure 1 shows the full outcome distribution. Graduates form the largest class, followed by dropouts and enrolled students. The enrolled class is analytically important but operationally ambiguous because it may represent delayed completion, continued registration, or risk that has not yet materialized.

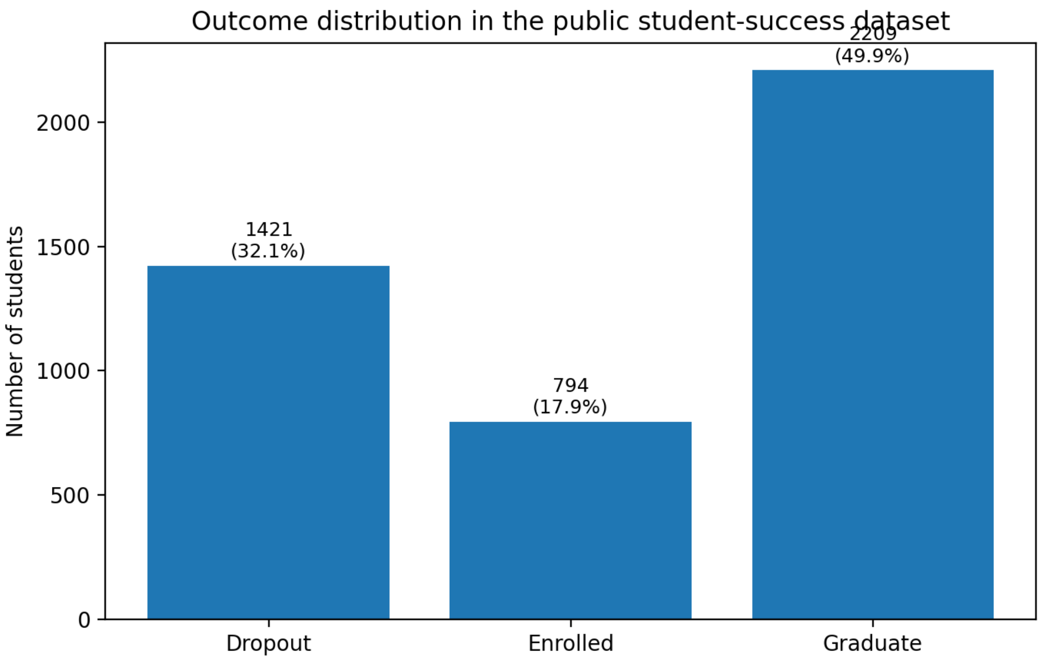


Figure 1: Outcome distribution in the public student-success dataset.

Figure 2 provides a four-panel view of the analysis. The first panel presents the decisive outcome cohort, the second compares hold-out accuracy, the third shows the operational error types for the governed logistic model, and the fourth maps risk intensity into intervention bands. The figure is placed directly after the corresponding explanation so that the descriptive and analytical evidence remain connected.

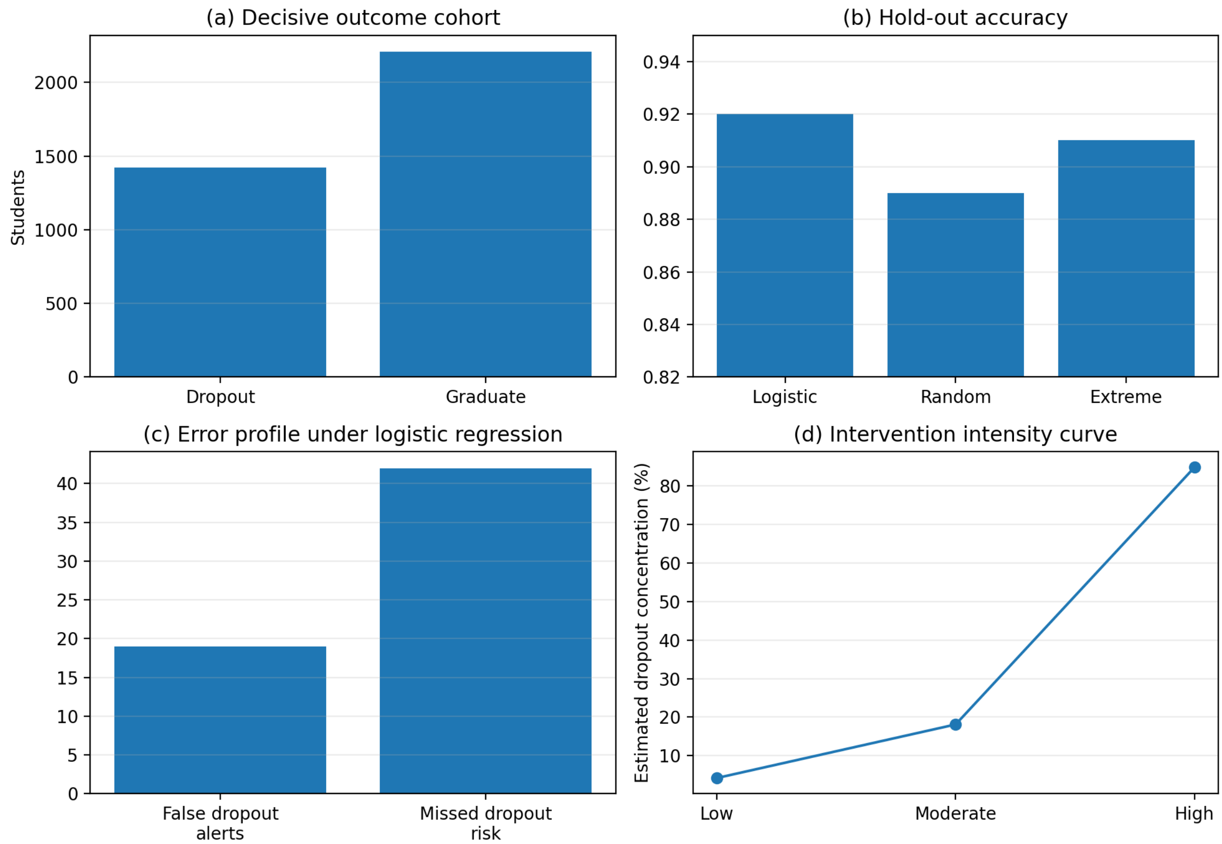


Figure 2: Four-panel evidence map for the proposed student-success analytics model.

Table 2 reports the held-out predictive results. Logistic regression achieved the strongest overall accuracy and weighted F1 in the decisive-outcome setting. This is important because the proposed framework values interpretability and operational accountability. The result suggests that more complex models do not automatically produce better business value when the objective is a governed early-warning process.

Table 2: Held-out performance for decisive student-success outcomes.

Model	Accuracy	Precision	Recall	Macro F1	Weighted F1
Logistic regression	0.92	0.92	0.92	0.91	0.92
Random forest	0.89	0.89	0.89	0.88	0.89
Extreme gradient boosting	0.91	0.91	0.91	0.91	0.91

Figure 3 compares the predictive performance of the three tested models. The relatively small performance gap between logistic regression and the more flexible models strengthens the case for using an interpretable specification when the model is intended for advisor-facing institutional use.

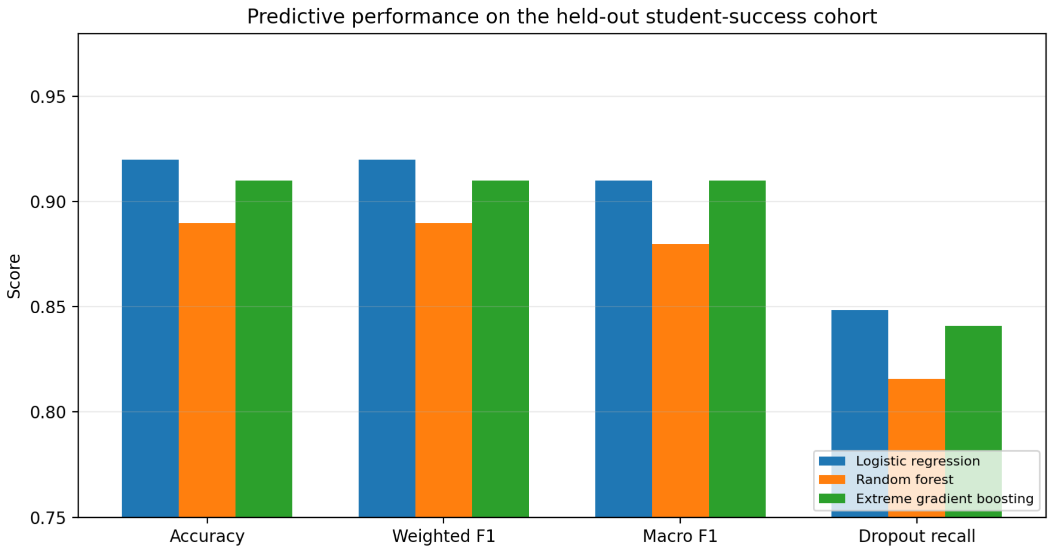
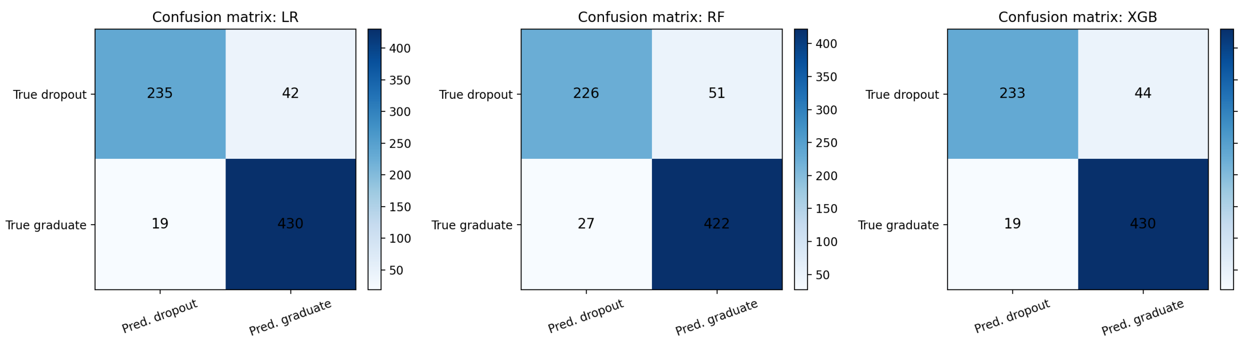


Figure 3: Comparative predictive performance across the three tested models.

The confusion matrices show a practical trade-off. Logistic regression produced 235 correctly detected dropout cases and 430 correctly detected graduate cases in the hold-out split. It missed 42 dropout cases and generated 19 false dropout alerts. In student-success management, these two errors have different costs. A missed dropout case may mean a lost opportunity for timely support, whereas a false alert may consume advising resources or create unnecessary concern.



(a) Logistic regression **(b)** Random forest **(c)** XGBoost

Figure 4: Confusion matrices for the decisive-outcome experiment.

Figure 5 translates predictive signals into an intervention priority map. The strongest operational signals are curricular-unit approval, semester grade, tuition status, and debtor status. This result is consistent with the broader literature that academic progression variables are usually stronger predictors than static demographic variables, but the framework still requires human review before high-impact intervention.

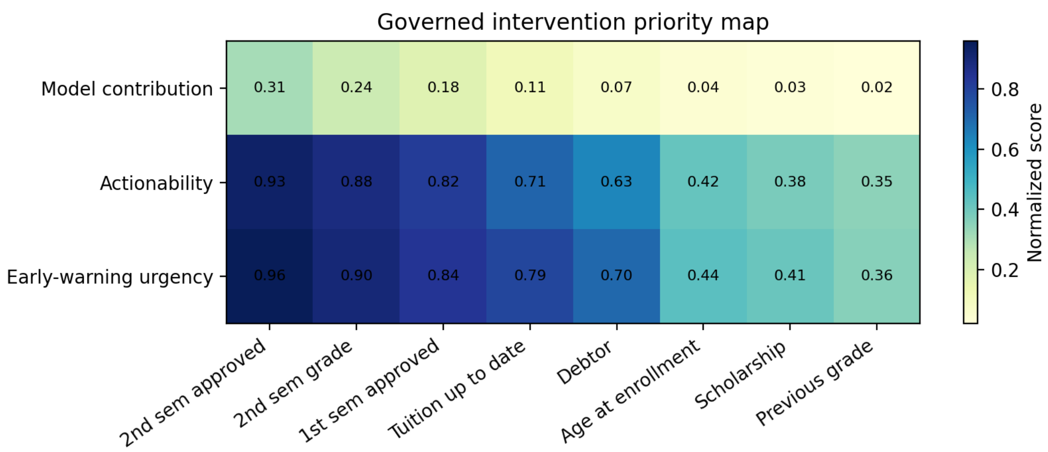


Figure 5: Governed intervention priority map based on predictive contribution, actionability, and urgency.

Figure 6 presents tri-class sensitivity using benchmark values reported for the same dataset. The enrolled class is harder to predict than dropout or graduate outcomes. This supports the decision to separate decisive outcome prediction from enrolment-state monitoring in the proposed model.

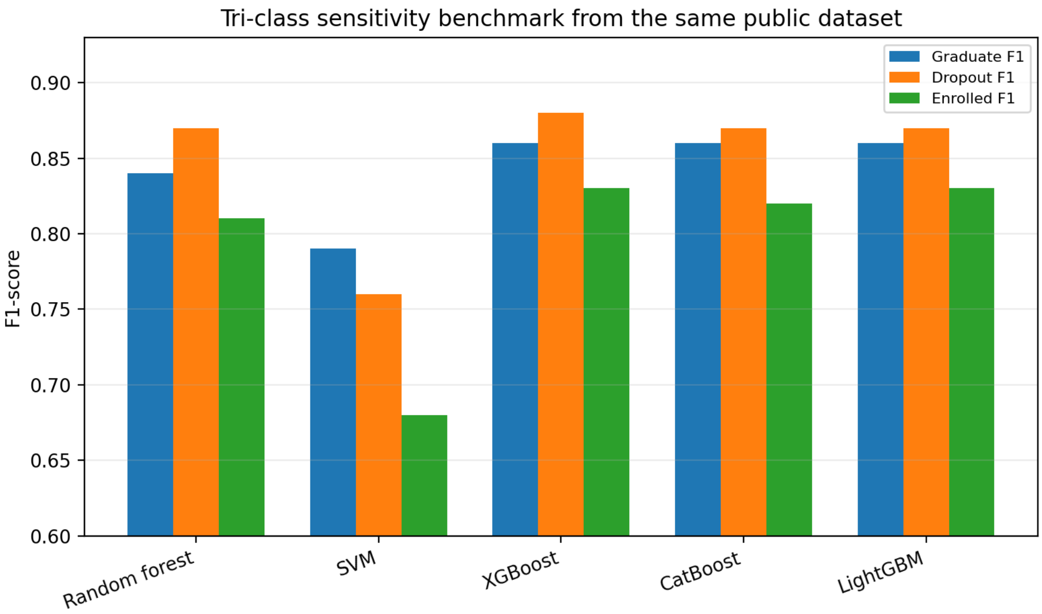


Figure 6: Tri-class sensitivity benchmark for the public student-success dataset.

6. Discussion

The findings offer five implications for education technology management. First, interpretable models can be competitive when the purpose is early warning rather than leaderboard performance. The logistic model delivered strong held-out results while preserving a clear explanation path for advisors. This matters because student-success decisions are social and institutional decisions, not only statistical outputs.

Second, the enrolled class requires a different managerial logic. Treating enrolled students as either failures or successes can distort the business question. A digital student-success unit should monitor enrolled students as a transitional portfolio and use additional time-sensitive indicators, such as LMS activity, course attempts, and advising history, before escalating risk.

Third, intervention design must be separated from prediction. The model identifies risk, but the institution decides what support is appropriate. Academic progression indicators may point to tutoring, study planning, or course redesign. Financial-status indicators may require fee-policy review or student-support coordination. Demographic variables should be handled carefully and should not become a basis for stigmatizing students.

Fourth, privacy and governance are not secondary requirements. Student-success systems can expose sensitive academic, socioeconomic, and personal information. The proposed governance layer ensures that risk scores are not released without explainability and policy approval. This aligns with recent privacy concerns in student-success information systems [15].

Fifth, predictive validity should be interpreted together with actionability and institutional trust. A technically accurate system that cannot be explained or acted upon will have limited value. Conversely, a transparent system with moderately high performance can be more valuable when it leads to consistent and timely student support. The results therefore support a business-oriented education technology view in which analytics becomes a managed service capability rather than a disconnected machine-learning experiment.

7. Conclusion

This paper presented a governed early-warning analytics model for student success in digital higher education. Using a public student-success dataset, the study showed that a transparent logistic model can perform competitively against more complex models while offering clearer value for academic advising and institutional decision-making. The proposed framework extends standard dropout prediction by adding intervention-priority and governance layers. The central conclusion is that student-success analytics should be designed as a responsible business capability: accurate enough to detect risk, interpretable enough to support action, and governed enough to protect students.

Future work should integrate longitudinal LMS logs, test model drift across academic years, compare fairness-aware thresholds, and evaluate whether recommended interventions improve retention outcomes in practice. The strongest next step is not only to predict which students are at risk, but to measure which forms of support help them succeed.

References

[1] V. Realinho, M. Vieira Martins, J. Machado, and L. Baptista, “Predict Students’ Dropout and Academic Success,” UCI Machine Learning Repository, 2021, doi: 10.24432/C5MC89.

[2] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, “Predicting student dropout and academic success,” *Data*, vol. 7, no. 11, p. 146, 2022, doi: 10.3390/data7110146.

[3] D. Ifenthaler and J. Y.-K. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Educational Technology Research and Development*, vol. 68, no. 4, pp. 1961–1990, 2020, doi: 10.1007/s11423-020-09788-z.

[4] A. Atif, D. Richards, D. Liu, and A. A. Bilgin, "Perceived benefits and barriers of a prototype early alert system to detect engagement and support at-risk students: The teacher perspective," *Computers & Education*, vol. 156, p. 103954, 2020, doi: 10.1016/j.compedu.2020.103954.

[5] J. Knobbout and E. Van Der Stappen, "Where is the learning in learning analytics? A systematic literature review on the operationalization of learning-related constructs in the evaluation of learning analytics interventions," *IEEE Transactions on Learning Technologies*, vol. 13, no. 3, pp. 631–645, 2020, doi: 10.1109/TLT.2020.2999970.

[6] M. Afzaal, J. Nouri, A. Zia, P. Papapetrou, U. Fors, Y. Wu, X. Li, and R. Weegar, "Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation," *Frontiers in Artificial Intelligence*, vol. 4, p. 723447, 2021, doi: 10.3389/frai.2021.723447.

[7] D. Al-Shaikhli, L. Jin, A. Porter, and A. Tarczynski, "Visualizing weekly learning outcomes and the intention to continue using a learning management system: The role of cognitive absorption and perceived learning self-regulation," *Education and Information Technologies*, vol. 27, no. 3, pp. 2909–2937, 2022, doi: 10.1007/s10639-021-10703-z.

[8] R. Jayashanka, E. Hettiarachchi, and K. P. Hewagamage, "Technology enhanced learning analytics dashboard in higher education," *The Electronic Journal of e-Learning*, vol. 20, no. 2, pp. 151–170, 2022, doi: 10.34190/ejel.20.2.2189.

[9] F. G. Karaoglan Yilmaz, "The effect of learning analytics assisted recommendations and guidance feedback on students' metacognitive awareness and academic achievements," *Journal of Computing in Higher Education*, vol. 34, no. 2, pp. 396–415, 2022, doi: 10.1007/s12528-021-09304-z.

[10] F. G. Karaoglan Yilmaz, "Utilizing learning analytics to support students' academic self-efficacy and problem-solving skills," *The Asia-Pacific Education Researcher*, vol. 31, no. 2, pp. 175–191, 2022, doi: 10.1007/s40299-020-00548-4.

[11] F. G. Karaoglan Yilmaz and R. Yilmaz, "Learning analytics intervention improves students' engagement in online learning," *Technology, Knowledge and Learning*, vol. 27, no. 2, pp. 449–460, 2022, doi: 10.1007/s10758-021-09547-w.

[12] T. K. Tan and L. Samavedham, "The learning process matters: A sequence analysis perspective of examining procrastination using learning management system," *Computers and Education Open*, vol. 3, p. 100112, 2022, doi: 10.1016/j.caeo.2022.100112.

[13] H. Lu, K. Huang, and T. E. Johnson, "The effects of different interventions on self-regulated learning of preservice teachers in a blended academic course," *Computers & Education*, vol. 181, p. 104442, 2022, doi: 10.1016/j.compedu.2022.104442.

[14] S. Kim, E. Choi, Y.-K. Jun, and S. Lee, "Student dropout prediction for university with high precision and recall," *Applied Sciences*, vol. 13, no. 10, p. 6275, 2023, doi: 10.3390/app13106275.

[15] M. Francis, M. Avoseh, K. Card, L. Newland, and K. Streff, "Student privacy and learning analytics: Investigating the application of privacy within a student success information system in higher education," *Journal of Learning Analytics*, vol. 10, no. 3, pp. 102–114, 2023, doi: 10.18608/jla.2023.7975.

[16] M. Mukred, U. A. Mokhtar, B. Hawash, H. AlSalman, and M. Zohaib, "The adoption and use of learning analytics tools to improve decision making in higher learning institutions: An extension of technology acceptance model," *Heliyon*, vol. 10, no. 4, p. e26315, 2024, doi: 10.1016/j.heliyon.2024.e26315.

[17] Z. Pan, L. Biegley, A. Taylor, and H. Zheng, "A systematic review of learning analytics: Incorporated instructional interventions on learning management systems," *Journal of Learning Analytics*, vol. 11, no. 2, pp. 52–72, 2024, doi: 10.18608/jla.2023.8093.

[18] M. Vaarma and H. Li, "Predicting student dropouts with machine learning: An empirical study in Finnish higher education," *Technology in Society*, vol. 76, p. 102474, 2024, doi: 10.1016/j.techsoc.2024.102474.

[19] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study," *Discover Artificial Intelligence*, vol. 4, p. 2, 2024, doi: 10.1007/s44163-023-00079-z.

[20] O. Goren, L. Cohen, and A. Rubinstein, "Early prediction of student dropout in higher education using machine learning models," in *Proceedings of the 17th International Conference on Educational Data Mining*, 2024, pp. 349–359, doi: 10.5281/zenodo.12729834.

[21] M. A. Hassan, A. H. Muse, and S. Nadarajah, "Predicting student dropout rates using supervised machine learning: Insights from the 2022 National Education Accessibility Survey in Somaliland," *Applied Sciences*, vol. 14, no. 17, p. 7593, 2024, doi: 10.3390/app14177593.