Creative Education, 2025, 16(8), 1107-1119
https://www.scirp.org/journal/ce
ISSN Online: 2151-4771
ISSN Print: 2151-4755

Predicting Student Success through Learning Analytics: A Comparative Modeling Approach

Artemis Rigou[1], Foteini Kyriazi[1], Dimitrios Thomakos[2]

[1]Department of Agribusiness & Supply Chain Management, Agricultural University of Athens, Thiva, Greece
[2]Department of Business Administration, National and Kapodistrian University of Athens, Athens, Greece
Email: artemisrigou@aua.gr, fkyriazi@aua.gr, dthomakos@ba.uoa.gr

How to cite this paper: Rigou, A., Kyriazi, F., & Thomakos, D. (2025). Predicting Student Success through Learning Analytics: A Comparative Modeling Approach. *Creative Education, 16,* 1107-1119.
https://doi.org/10.4236/ce.2025.168069

Received: July 11, 2025
Accepted: August 12, 2025
Published: August 15, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

In this paper, we explore the application of predictive modeling within the field of Learning Analytics (LA) to forecast student academic success in higher education. Utilizing the Open University Learning Analytics Dataset (OULAD), we integrate student demographic, educational, and assessment data to build a dataset suitable for supervised learning. Two models are employed: logistic regression, chosen for its interpretability, and Random Forest, selected for its capacity to capture complex, non-linear relationships. Our target variable is whether a student passes a course module. The analysis reveals that performance in early assessments is the most influential predictor, followed by prior education level and age group. The Random Forest model consistently outperforms logistic regression across all performance metrics, including accuracy, precision, and recall. These results emphasize the potential of machine learning to support early identification of at-risk students, guiding timely interventions. We conclude by discussing the policy implications of our findings for institutional strategies aimed at improving student retention and academic outcomes.

Keywords

Forecasting, Learning Analytics, Prediction, Higher Education

1. Introduction

Forecasting student academic performance is a central task in the field of Learning Analytics (LA), a discipline that merges educational theory with data science to improve teaching, learning, and student success. As digital learning environments proliferate, institutions are amassing vast amounts of behavioral and performance-related data. These data offer unprecedented opportunities to identify students at risk of failure and implement timely interventions (Long & Siemens, 2014;

Gourna et al., 2024). Predictive modeling within LA focuses on leveraging machine learning techniques to forecast important educational outcomes, including course completion, assignment success and graduation rates. Among these, predicting whether a student will pass or fail a course is particularly valuable for course designers, tutors, and administrators, enabling them to allocate resources more effectively and enhance the quality of instruction (Baker & Inventado, 2014).

In this paper, we utilize the Open University Learning Analytics Dataset (OULAD), a publicly available dataset that contains rich information on student demographics, academic history, and assessment results. Our dataset is a cross-sectional panel data organized around student-module interactions. Our goal is to construct models that predict whether a student will pass a module based on demographic and performance features available early in the course. We apply both logistic regression and random forest classification to model the binary outcome of course success. Logistic regression provides a transparent and interpretable baseline model, while random forests offer a robust non-linear alternative capable of capturing complex interactions among features. The performance of both models is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrices. Additionally, we examine feature importances to understand the primary drivers of predictive performance.

By systematically comparing linear and ensemble methods on real educational data, this paper contributes to the growing literature on predictive learning analytics. It also highlights how structured institutional data can be harnessed to support early-warning systems and inform policy decisions within higher education.

The remainder of the paper is organized as follows: Section 2 presents a brief literature review on learning analytics; Section 3 describes the data and methodology; Section 4 reports the results; Section 5 provides a discussion; Finally, the conclusions and policy implications are highlighted in Section 6.

2. Brief Literature Review

Learning Analytics (LA) has emerged as a pivotal field in educational research, focusing on leveraging data to improve teaching effectiveness and student learning outcomes. Siemens (2019) emphasizes the power of LA in open and distance learning, noting that access to granular, learner-generated data allows algorithmic insights into behavior and engagement. The integration of LAs into educational contexts has the potential to transform instruction through early intervention and personalized support strategies.

A foundational study by Papamitsiou and Economides (2014) provides a systematic literature review of empirical work in Learning Analytics and Educational Data Mining (EDM), categorizing applications into four strategic directions: prediction, intervention, personalization, and reflection. Their analysis underscores LA's capacity to inform adaptive learning pathways and improve learner outcomes through continuous feedback loops. Nguyen et al. (2020) developed a predictive model utilizing interaction data from a Moodle-based LMS to forecast learner out-

comes in blended learning environments. Their model, incorporating data mining and machine learning techniques, demonstrated an accuracy exceeding 50%, highlighting the feasibility of data-driven predictions in online education settings.

In contrast, Rincon-Flores et al. (2022) explored the use of K-Nearest-Neighbor and Random Forest algorithms to predict student profiles in Physics courses. Their findings revealed that prediction accuracy improved as the models were exposed to more data over time, supporting the use of incremental learning and adaptive algorithms in academic environments. In a similar context, Alhammad (2023) applied LA to English as a Foreign Language (EFL) classrooms, revealing a strong correlation between engagement with video content and final exam scores. Her findings suggest that instructional design, especially the incorporation of video, plays a crucial role in predictive modeling and learner success.

Mubarak et al. (2020) took a deep learning approach, employing Long Short-Term Memory (LSTM) networks to analyze clickstream data from MOOC video lectures. Their model achieved up to 93% accuracy in predicting learner performance, showcasing the strength of temporal modeling in high-volume educational data.[1] Collectively, these studies demonstrate the maturity of LA as a field that incorporates a variety of methodologies—from interpretable models like logistic regression to black-box approaches like deep learning. Importantly, they underline the need for timely data, ethical considerations, and a strong connection to pedagogical frameworks to ensure meaningful application.

The practice of learning analytics has caused widespread changes in several areas of higher education, including the capacity to identify students who are at risk (Kotsiantis et al., 2013; Kulik et al., 1990; Bañeres et al., 2020), monitor learners' advancement (Libbrecht et al., 2012; Gupta & Yadav, 2023), predicting each student's individual learning needs(Krumm et al., 2014; Arnold & Pistilli, 2012; Beck and Davidson, 2001), revealing potential determinants of academic performance (Greller & Drachsler, 2012) and learners' behaviors (Bousbia & Belamri, 2014). Moreover, Gašević et al. (2016) found that the counting and tracking of online learning activities of learners using learning tools are correlated with learners' academic performance, but Rogers et al. (2016) highlights that Learning analytics sometimes provide insufficient summative data about students' academic performance and learning process. As far as the geographic extend of data, Baker and Yacef (2009) found that the use of educational data mining is dominant in some regions and with relatively limited usage in other regions, fact that may hinder the adoption of learning analytics as its potential benefits are inhibited.

In the meantime, Wang (2016) points out the technical and analytical challenges posed by enormous amounts of unstructured data when using a scalable approach to improve students' learning and at the same time Giacumo and Bremen (2016) apply that there is little evidence for the effective use of big data analytic

[1]See also Guerard et al. (2023); Kyriazi (2024), which they apply advanced methodologies such as distance-based nearest neighbor forecasting and prescriptive techniques, offering robust frameworks that can provide valuable cross-disciplinary insights for predictive modeling in educational contexts, particularly in the areas of model stability, validation, and decision support.

methodologies. Similarly, Viberg et al. (2018) demonstrated that learning analytics tools have had little impact on students' learning outcomes (9%) and on learning support and teaching (35%). However, analysis of existing learning analytics evidence indicates a shift towards a deeper understanding of students' learning experiences in recent years.

In the context of our research, which utilizes the Open University Learning Analytics Dataset (OULAD), these works guide the selection of predictive features and model architectures. They also reinforce the value of using models such as logistic regression for transparency and Random Forests for capturing complex, non-linear relationships. The inclusion of such literature provides both validation and theoretical grounding for the predictive strategies explored in this paper.

3. Data and Methodology

3.1. Data

Our dataset stems from the Open University Learning Analytics Dataset (OULAD), a publicly available and anonymized collection of data released by the UK's Open University. It includes detailed records from several course presentations between 2013 and 2014, making it particularly well-suited for research in educational data mining and predictive learning analytics. The dataset comprises observations from multiple academic presentations and is best characterized as cross-sectional panel data, structured around individual students and their enrollment in specific course modules, rather than a continuous time series.

For the purposes of our analysis, we focus on two core components of the OULAD dataset: The student information and student assessment records. The former provides demographic and educational background information, including variables such as gender, age band, highest prior education, disability status, index of multiple deprivation (IMD band), number of previous attempts, studied credits, and the student's final result (Pass, Fail, Withdrawn, or Distinction). The latter contains assessment-level data, capturing individual assignment and exam scores for each student. By merging these tables, we constructed a dataset suitable for predictive modeling. Our target variable was binary—whether a student passed or not—and predictors included both demographic attributes and a calculated average assessment score. For this study, early assessments are defined as the first two assessments submitted by each student. These scores were averaged to create a feature capturing early academic performance, allowing timely risk identification. This combined structure allowed us to train classification models to identify the likelihood of academic success. This combined structure allowed us to train classification models to identify the likelihood of academic success.

Before modeling, we applied several preprocessing steps to prepare the data. All categorical variables were label encoded to transform them into a numerical format suitable for classification models. We also handled missing data by performing complete-case analysis, removing any records with missing values. Since our primary classifier was a Random Forest, we did not apply normalization, given

the model's invariance to feature scaling. However, we ensured a randomized train/test split (80/20) to evaluate model performance on unseen data. The target variable was binarized to distinguish between "Pass" and all other outcomes. These steps resulted in a clean and structured dataset appropriate for predictive modeling in learning analytics.

3.2. Methodology

We employ a supervised machine learning framework to predict student academic success, defined as achieving a final course result of "Pass." The process consists of five main stages: data integration, preprocessing, feature engineering, model selection, and performance evaluation. We use two primary data sources from the Open University Learning Analytics Dataset (OULAD), corresponding to student background and assessment performance. These datasets were merged using a unique student identifier to combine demographic, educational, and performance information. To ensure consistency and avoid issues caused by missing values, we performed complete-case analysis, retaining only records with full information across all relevant variables.

As part of feature engineering, we created a new numeric variable representing each student's average score across all their assessments. Categorical variables such as gender, age band, highest education, disability status, and IMD band were converted into numerical values using label encoding. The target variable was binary, indicating whether a student passed the module (1 = Pass, 0 = otherwise). No normalization or scaling was applied, as our main classifier—Random Forest—is inherently insensitive to feature magnitude.

We implemented two classification models. First, a logistic regression model served as a benchmark. This model estimates the probability that a student will pass as:

$$P\left(y=1\middle|x\right)=\frac{1}{1+\mathrm{e}^{-x\top\beta}} \tag{1}$$

where x is the feature vector and β is the vector of estimated coefficients.

Parameters are learned via maximum likelihood estimation, maximizing the log-likelihood function:

$$\log L\left(\beta\right)=\sum_{i=1}^{n}\left[y_i\log\left(P_i\right)+\left(1-y_i\right)\log\left(1-P_i\right)\right] \tag{2}$$

where P_i is the predicted probability of student i passing.

The second model was a Random Forest classifier, an ensemble of decision trees trained on bootstrap samples. Each tree performs recursive binary splits on features selected from a random subset. The split criterion minimizes node impurity, measured using the Gini index:

$$Gini\left(t\right)=1-\sum_{c=1}^{C}p_c^2 \tag{3}$$

where p_c is the proportion of class c samples in node t. Final predictions are made

via majority vote across trees.

For our experiments, we used scikit-learn's default hyperparameters:

- n_estimators = 100
- max_depth = None
- min_samples_split = 2
- min_samples_leaf = 1
- max_features = 'sqrt'

We chose to retain the default hyperparameters in scikit-learn to establish a consistent and reproducible baseline. While this choice may limit model performance, it ensures a fair comparison with logistic regression and reflects typical usage. Future work could apply systematic hyperparameter tuning to explore performance improvements. These settings allowed the model to grow complex trees while benefiting from ensemble averaging to reduce variance and prevent overfitting. The dataset was randomly split into training (80%) and testing (20%) sets.

Model performance was evaluated using classification accuracy, precision, recall, and F1-score. A confusion matrix provided further insight into false positives and false negatives. Additionally, we extracted feature importance scores from the Random Forest model to identify key predictors of student success.

4. Results

The evaluation of our predictive models focused on several standard classification metrics, including accuracy, precision, recall, and F1-score, computed using the test set that was held out during training. These metrics provide insights into both the overall performance and the behavior of the models with respect to different types of classification errors.

The logistic regression model demonstrated balanced but limited performance, achieving moderate precision and recall across both classes. This result is expected given the model's linear assumptions and limited ability to capture complex patterns or interactions among input features. Despite this, its interpretability remains a strength, particularly in identifying the direction and strength of influence of individual predictors.

The Random Forest classifier outperformed logistic regression across several metrics. Its ensemble structure, based on bagging and random feature selection, enabled it to handle the non-linearities and heterogeneous patterns present in the student population data. The confusion matrix revealed that the Random Forest model was more effective in correctly identifying students who passed their modules (true positives), while also reducing false positive predictions compared to logistic regression.

Table 1 presents the set of features used in the predictive modeling of student success. These variables were extracted from the OULAD dataset and encompass both demographic characteristics (such as gender, age group, and socio-economic status) and academic information (including assessment scores and credit load). Each feature is categorized by type—categorical, numerical, or binary—to reflect its role in the machine learning models. This structured summary provides clarity on

the data inputs that underpin the logistic regression and random forest classifiers.

The descriptive analysis presented in Table 2 and Table 3 offers valuable insights into the demographic and academic composition of the student population used in our predictive modeling. The distribution of key variables, such as age band, highest education level, and average assessment score, reveals important trends that are directly aligned with the performance of our models. For instance, the prevalence of adult learners aged 35 - 55, as well as students holding A-level or higher qualifications, underscores the importance of educational background and age as predictors of academic success. These variables were found to be among the most influential in the Random Forest model's feature importance analysis. Moreover, the summary statistics in Table 2 show substantial variability in assessment scores and credits studied, further justifying their inclusion as predictors. The strong predictive power of the average assessment score, in particular, is supported by its wide distribution and its clear relationship with course outcomes. The patterns observed in the categorical frequency distributions (Table 3) also suggest that certain demographic groups, such as students without declared disabilities and those from mid-to-high IMD bands, are more represented among those who succeed, a factor that may enhance model sensitivity.

Table 1. Description of features used in predictive modeling.

Feature Name	Description	Type
gender	Gender of the student (M/F)	Categorical
age_band	Age group (e.g., 0 - 35, 35 - 55)	Categorical
highest_education	Highest prior education level	Categorical
imd_band	Socio-economic band (IMD)	Categorical
disability	Disability status (Yes/No)	Categorical
region	Region of residence	Categorical
studied_credits	Number of credits studied	Numerical
final_score	Final numeric score in assessment	Numerical
average_assessment_score	Average score across assessments	Numerical
Pass	Target variable: Pass (1) or Fail (0)	Binary

Note: This table summarizes the student-level features drawn from the OULAD dataset. These variables were selected for their relevance to demographic and academic engagement characteristics.

Table 2. Summary statistics of studied credits.

	Count	Mean	Std Dev	Min	25%	50%	75%	Max
studied credits	32.593	59.79	59.52	0	30	60	90	120

Note: This table summarizes the distribution of the number of credits studied by each student in the OULAD dataset. The 25%, 50%, and 75% values correspond to the first, second (median), and third quartiles, respectively.

Table 3. Frequencies of key categorical variables.

Variable	Category	Count
age_band	0 - 35	10.441
	35 - 55	13.989
	55+	8.163
gender	Female (F)	17.689
	Male (M)	14.904
highest_education	Lower Than a Level	3.385
	A Level or Equivalent	10.513
	HE Qualification	13.843
	Postgraduate Qualification	3.852
disability	No	29.613
	Yes	2.990
imd_band	0% - 10%	470
	10% - 20%	5.162
	20% - 30%	8.867
	30% - 40%	8.254
	40% - 50%	7.840
final_result	Distinction	5.787
	Pass	18.987
	Withdrawn	3.866
	Fail	1.953

Note: This table presents the frequencies for selected categorical variables in the dataset. Each variable reflects an aspect of student demographics or academic status used in the predictive models.

5. Discussion

Taken together, these descriptive findings not only validate our feature selection process but also provide context for interpreting the relative strengths of the Random Forest and logistic regression models. The demographic and academic characteristics of the sample appear well-suited to predictive modeling, offering structured variation that the models can effectively exploit to distinguish between successful and at risk students.

Continuing with **Table 4**, which presents the classification report for the Random Forest model, we observe that the model achieved a precision of 0.56 and recall of 0.52 for the 'Pass' class (Class 1), indicating moderate ability in identifying successful students. For the 'Not Pass' class (Class 0), performance was slightly better, with precision of 0.62 and recall of 0.65. The overall accuracy and macro average F1-score both stood at 0.59, reflecting modest but balanced performance across both classes. Precisely, the confusion matrix (**Figure 1**) visually confirms these results: A reasonable proportion of students who passed were correctly identified, although misclassifications still occurred. The number of students who failed but were predicted to pass highlights a potential limitation in sensitivity,

particularly for the positive class. Feature importance analysis (Figure 2) clearly shows that the most influential variable in predicting student success was the average assessment score, followed by demographic factors such as age band and highest education level. While the overall accuracy of 0.59 may appear moderate, this level of predictive power is meaningful in the context of early-warning systems. Identifying at-risk students with over 50% reliability enables institutions to initiate light, preventive interventions (e.g., automated messages or tutoring offers), with minimal cost even in the case of false positives. These results emphasize the central role of academic engagement and past performance in shaping educational outcomes. Demographic characteristics, while relevant, are secondary to behavioral indicators of success.

Table 4. Classification report for random forest model.

	Precision	Recall	F1-score
Class 0	0.62	0.65	0.63
Class 1	0.56	0.52	0.54
Accuracy	0.59	0.59	0.59
Macro avg	0.59	0.59	0.59

Note: The Random Forest classifier demonstrates moderately balanced predictive power, showing slightly stronger performance in predicting non-pass outcomes, as is common in imbalanced datasets.

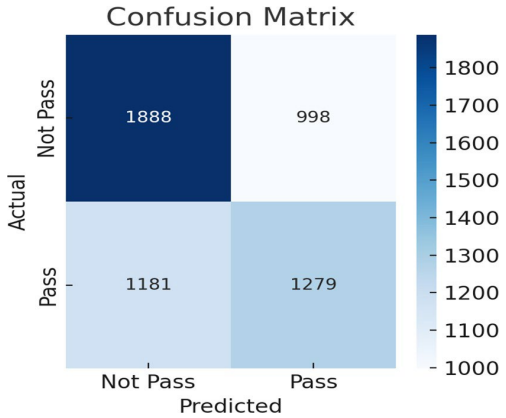

Figure 1. Confusion matrix for random forest classifier.

Note: The Random Forest model shows strong performance in predicting successful students (true positives), with relatively low false positives. However, there are still some misclassifications among students who failed.

Overall, these findings support the conclusion that the Random Forest model offers a flexible and generalizable approach for early prediction of student success. While it improves on logistic regression, future work should explore enhancing prediction for the minority class, potentially through class weighting, resampling strategies, or the inclusion of richer behavioral data such as student interaction with virtual learning environments (VLEs).

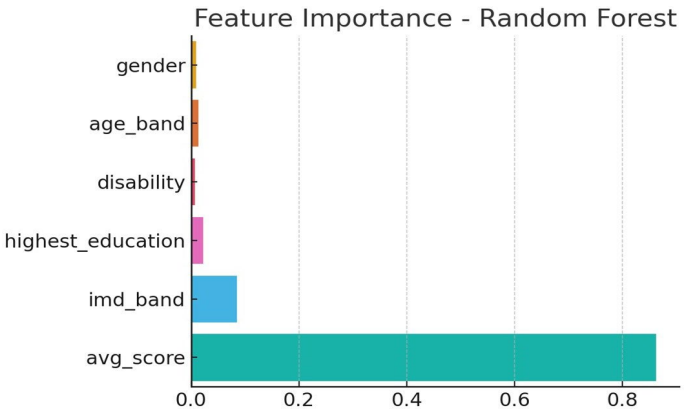

Figure 2. Feature importance ranking (random forest).

Note: The average assessment score is by far the most significant predictor, followed by highest education and age band. Demographic variables show smaller but non-negligible influence.

In contrast to the Random Forest approach, the logistic regression model produced more moderate results. As a generalized linear model, logistic regression attempts to draw a linear decision boundary between the two outcome classes. While this model offers interpretability and computational efficiency, it is inherently limited in capturing non-linear relationships present in educational data. As shown in Table 5, logistic regression achieved a precision of 0.53 and recall of 0.45 for the 'Pass' class. For the 'Not Pass' class, the precision and recall were 0.58 and 0.66, respectively. These values are slightly lower than those of the Random Forest model, particularly in terms of recall for successful students. The confusion matrix (Figure 3) further underscores these limitations, showing that logistic regression misclassified a larger number of failing students as passing—highlighting its reduced sensitivity to the minority class and the challenges of applying linear models to complex educational data.

Table 5. Logistic regression classification report.

	Precision	Recall	F1-score
Class 0	0.58	0.66	0.62
Class 1	0.53	0.45	0.49
Accuracy	0.56	0.56	0.56
Macro avg	0.56	0.55	0.55

Note: Logistic regression serves as a baseline model, with slightly lower overall accuracy and notably reduced recall for the positive class. Its advantage lies in interpretability rather than predictive strength.

In summary, logistic regression serves as a useful baseline model due to its simplicity and transparency. However, its predictive performance on this dataset is inferior to that of ensemble methods, making it better suited for applications where interpretability is prioritized over predictive accuracy.

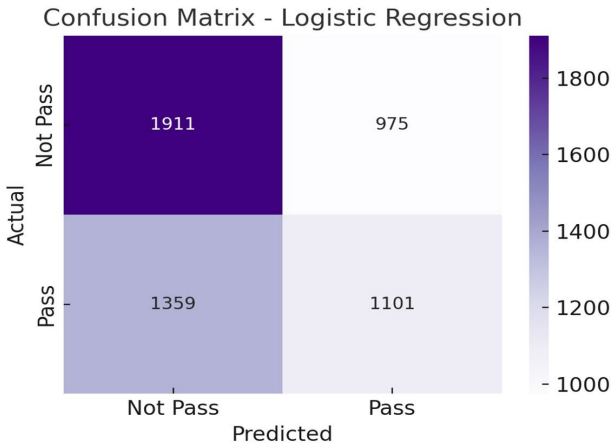

Figure 3. Confusion matrix for logistic regression.

Note: The logistic regression model misclassifies a significant portion of students who failed, confirming its limited capacity to separate the classes effectively in imbalanced data settings.

6. Conclusion and Policy Implications

In this paper, we demonstrate the potential of predictive modeling techniques to identify students at risk of failing academic modules, using structured institutional data from the Open University Learning Analytics Dataset (OULAD). Through a comparative analysis of logistic regression and random forest models, we show that ensemble-based approaches yield significantly stronger predictive performance, particularly in accurately identifying successful students. Assessment performance emerged as the most critical predictor, followed by educational background and age.

These findings highlight the important role of learning analytics in informing evidence-based policy and educational intervention. Institutions can leverage similar predictive models to implement early-warning systems, allocate support resources more effectively, and tailor academic interventions based on data-driven insights. In particular, educators and administrators may consider continuous monitoring of student progress and use model outputs to initiate timely outreach or instructional adjustments.

From a broader policy perspective, the results support the integration of predictive analytics into higher education strategy. Institutions should invest in robust data infrastructure and promote a culture of ethical analytics use—one that balances predictive power with transparency, fairness, and student consent. Future studies should consider incorporating behavioral engagement data and temporal learning patterns, which may further enhance prediction accuracy and provide deeper insights into student learning trajectories.

There are several limitations to note. First, our use of complete-case analysis may introduce bias if the excluded data are not missing completely at random. Second, using default hyperparameters for the Random Forest classifier means that results could be improved through tuning. Third, the dataset originates from a single institution (The Open University), which may limit the generalizability of

our findings to other educational contexts without further validation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Alhammad, A. I. (2023). Learning Analytics in the EFL Classroom: Technology as a Forecasting Tool. *Research Journal in Advanced Humanities, 4,* 97-108. https://doi.org/10.58256/rjah.v4i3.1194

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270). ACM. https://doi.org/10.1145/2330601.2330666

Baker, R. S. J., & Inventado, P. S. (2014). Chapter X: Educational Data Mining and Learning Analytics. *Computer Science, 7,* 1-16.

Baker, R. S., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining, 1,* 3-17.

Bañeres, D., Rodríguez, M. E., Guerrero-Roldán, A. E., & Karadeniz, A. (2020). An Early Warning System to Detect At-Risk Students in Online Higher Education. *Applied Sciences, 10,* Article 4427. https://doi.org/10.3390/app10134427

Beck, H. P., & Davidson, W. D. (2001). Establishing an Early Warning System: Predicting Low Grades in College Students from Survey of Academic Orientations Scores. *Research in Higher Education, 42,* 709-723. https://doi.org/10.1023/a:1012253527960

Bousbia, N., & Belamri, I. (2014). Which Contribution Does EDM Provide to Computer-Based Learning Environments? In *Studies in Computational Intelligence* (pp. 3-28). Springer. https://doi.org/10.1007/978-3-319-02738-8_1

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning Analytics Should Not Promote One Size Fits All: The Effects of Instructional Conditions in Predicting Academic Success. *The Internet and Higher Education, 28,* 68-84. https://doi.org/10.1016/j.iheduc.2015.10.002

Giacumo, L. A., & Bremen, J. (2016). Emerging Evidence on the use of Big Data and Analytics in Workplace Learning: A Systematic Literature Review. *Quarterly Review of Distance Education, 17,* 21-38.

Gourna, S., Rigou, A., Kyriazi, F., & Marinagi, C. (2024). The Added Value of Learning Analytics in Higher Education. *International Journal of Education and Information Technologies, 18,* 133-142. https://doi.org/10.46300/9109.2024.18.13

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology and Society, 15,* 42-57.

Guerard, J., Thomakos, D., Kyriazi, F., & Mamais, K. (2023). On the Predictability of the DJIA and S&P500 Indices. *Wilmott Magazine, 129,* 1-84.

Gupta, O. J., & Yadav, S. (2023). Determinants in Advancement of Teaching and Learning in Higher Education: In Special Reference to Management Education. *The International Journal of Management Education, 21,* Article 100823. https://doi.org/10.1016/j.ijme.2023.100823

Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using Learning Analytics to Identify Successful Learners in a Blended Learning Course. *International Journal of*

Technology Enhanced Learning, 5, 133-150. https://doi.org/10.1504/ijtel.2013.059088

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering. In *Learning Analytics* (pp. 103-119). Springer. https://doi.org/10.1007/978-1-4614-3305-7_6

Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of Mastery Learning Programs: A Meta-Analysis. *Review of Educational Research, 60,* 265-299. https://doi.org/10.3102/00346543060002265

Kyriazi, F. (2024). The Prescriptive Nature of Market Timing and Predictive Portfolios. *IMA Journal of Management Mathematics, 36,* 323-338. https://doi.org/10.1093/imaman/dpae027

Libbrecht, P., Rebholz, S., Herding, D., Müller, W., & Tscheulin, F. (2012). Understanding the Learners' Actions When Using Mathematics Learning Tools. In *Lecture Notes in Computer Science* (pp. 111-126). Springer. https://doi.org/10.1007/978-3-642-31374-5_8

Long, P. D., & Siemens, G. (2014). Penetrating the Fog: Analytics in Learning and Education. *Tecnologie Didattiche, 22,* 132-137. https://doi.org/10.17471/2499-4324/195

Mubarak, A. A., Cao, H., & Ahmed, S. A. M. (2020). Predictive Learning Analytics Using Deep Learning Model in MOOCs' Courses Videos. *Education and Information Technologies, 26,* 371-392. https://doi.org/10.1007/s10639-020-10273-6

Nguyen, A., Gardner, L., & Sheridan, D. (2020). Data Analytics in Higher Education: An Integrated View. *Journal of Information Systems Education, 31,* 61-71. https://aisel.aisnet.org/jise/vol31/iss1/5

Papamitsiou, Z., & Economides, A. A. (2014). Temporal Learning Analytics for Adaptive Assessment. *Journal of Learning Analytics, 1,* 165-168. https://doi.org/10.18608/jla.2014.13.13

Rincon-Flores, E. G., Lopez-Camacho, E., Mena, J., & Olmos, O. (2022). Teaching through Learning Analytics: Predicting Student Learning Profiles in a Physics Course at a Higher Education Institution. *International Journal of Interactive Multimedia and Artificial Intelligence, 7,* 82-89. https://doi.org/10.9781/ijimai.2022.01.005

Rogers, T., Gašević, D., & Dawson, S. (2016). Learning Analytics and the Imperative for Theory-Driven Research. *The SAGE Handbook of E-learning Research,* 232-250. https://doi.org/10.4135/9781473955011.n12

Siemens, G. (2019). Learning Analytics and Open, Flexible, and Distance Learning. *Distance Education, 40,* 414-418. https://doi.org/10.1080/01587919.2019.1656153

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The Current Landscape of Learning Analytics in Higher Education. *Computers in Human Behavior, 89,* 98-110. https://doi.org/10.1016/j.chb.2018.07.027

Wang, Y. (2016). Big Opportunities and Big Concerns of Big Data in Education. *TechTrends, 60,* 381-384. https://doi.org/10.1007/s11528-016-0072-1