International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Efficient Data Analysis Model Integrating Blended Learning and Learner Engagement in Higher Education Institutions

[1]Kataru Anilkumar, [2]Venkateswararao Podile

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: As the digital age propels forward, the engagement of learners within these hybrid settings becomes increasingly critical to their academic success and overall retention rates. The effective integration of Educational Data Mining (EDM) and Learning Analytics (LA) emerges as a paramount strategy in understanding and optimizing learner engagement and academic progression. This study presents a pioneering model that integrates blended learning and learner engagement through the application of Educational Data Mining (EDM) and Learning Analytics (LA) in higher education institutions. By employing advanced data analytics techniques, including a hybrid boosting classifier, the research identifies critical factors influencing student academic progression and retention rates. The analysis of an extensive dataset covering various aspects of learner engagement—such as technology usage, instructor interaction, and feedback quality—reveals significant insights. These insights enable the prediction of student outcomes, offering a novel approach to enhance educational delivery and support mechanisms. The findings highlight the potential of machine learning models in transforming educational strategies and fostering a deeper understanding of student engagement within blended learning environments.

Keywords*: Data analysis, Learning analytics, Learners engagement, blended learning.*

1. Introduction

In the higher education, student academic progression stands as a pivotal element determining the success and retention rates within university programs. This progression denotes the movement of students from one educational level to another or their advancement towards program completion. However, the effectiveness of this progression is contingent upon various factors[1]. This study delves into identifying the influential parameters that significantly impact student performance and retention rates. A burgeoning body of research in Educational Data Mining (EDM) is dedicated to mitigating student attrition rates in universities. With the evolution of education delivery methods, traditional classroom teaching has transitioned towards more technologically integrated approaches. Many universities now offer programs online or in blended modes, capitalizing on the forefront role of technology. Nevertheless, online programs often grapple with high attrition rates due to the lack of robust monitoring mechanisms. Researchers are actively engaged in exploring solutions to this challenge by identifying influential factors that affect student retention [2]. Conversely, traditional classroom systems face a dearth of data compared to online education modes, where valuable insights can be derived from log file data. Thus, understanding and addressing the complex interplay between blended learning approaches and learner engagement variables are essential in fostering student success and program completion in higher education [3]. In the traditional educational environment, crucial data points related to student engagement and learning behaviors are often inaccessible, limiting further research and hindering the ability of institutions to make informed decisions. While some learning management systems (LMS) like Moodle and Blackboard are in use, they may not capture comprehensive data on various aspects of learning, such as objectives, actions, and participation. Consequently, educational institutions face challenges in accurately predicting graduation rates, placements, and overall student success [4]. To address these challenges, Educational Data Mining (EDM) and Learning Analytics (LA) techniques have emerged, enabling institutions to collect extensive data on student performance, motivation, and resilience. These methodologies employ Data Mining (DM) and Machine Learning (ML) models, categorized into supervised and unsupervised learning approaches, to predict student outcomes and identify factors influencing academic performance. For instance, Bayesian Profile Regression is utilized to pinpoint students at risk of dropping out based on their performance, motivation, and resilience. Additionally, the integration of DM in higher education has become increasingly critical, with EDM and LA methodologies offering innovative solutions to interaction-related challenges [5]. Through EDM, educational institutions can analyze vast datasets to extract meaningful patterns and inform decision-making

[1]*Research scholar, K.L.Business School , Koneru Lakshmiah Education Foundation, Vaddeswaram,kataruanilkumarsj@gmail.com,*
[2]*Professor , K.L.Business School , Koneru Lakshmiah Education Foundation, Vaddeswaram,vraopodile@gmail.com*

processes, akin to how Machine Learning is applied in other industries like healthcare and banking. Researchers have employed various algorithms, such as Decision Trees, Support Vector Machines, Random Forest, and Gradient Boosting, to classify students based on their academic outcomes and identify influential factors impacting their performance. The adoption of ensemble learners like Random Forest and Gradient Boosting has shown promising results in predicting student outcomes accurately, thereby highlighting the potential of ML in enhancing student engagement and academic success in blended learning environments [6].

Training data and Research questions

In the context of blended learning, learner engagement variables encompass a range of factors that influence how actively and effectively students participate in the learning process across both traditional face-to-face instruction and online learning environments. Here are some key learner engagement variables in blended learning [7-10]:Technology Usage: This variable measures the extent to which students utilize digital tools, platforms, and resources in their learning activities. It includes factors such as frequency of accessing online materials, participation in virtual discussions, and utilization of multimedia resources.

Instructor Interaction: This variable gauges the level of interaction and support provided by instructors in both physical and virtual settings. It includes factors such as responsiveness to student queries, provision of timely feedback on assignments, and facilitation of online discussions.

Social Learning Opportunities: This variable assesses the extent to which students engage in collaborative learning activities with peers, both in-person and online. It includes factors such as participation in group projects, peer review activities, and online discussions.

Feedback Quality: This variable evaluates the effectiveness and helpfulness of feedback provided to students by instructors and peers. It includes factors such as clarity of feedback, specificity of suggestions for improvement, and timeliness of feedback delivery.

Course Content Relevance: This variable measures the perceived relevance and alignment of course content with students' interests, goals, and real-world applications. It includes factors such as the incorporation of authentic tasks, case studies, and real-life examples into the curriculum.

Autonomy and Self-Regulation: This variable assesses students' ability to take ownership of their learning and regulate their learning behaviors effectively. It includes factors such as goal setting, time management skills, and self-assessment practices.

Motivation: This variable evaluates students' level of interest, enthusiasm, and persistence in engaging with course materials and completing learning tasks. It includes factors such as intrinsic motivation, extrinsic incentives, and perceived competence.

Access to Support Resources: This variable considers the availability and accessibility of support services and resources to assist students in their learning journey. It includes factors such as access to tutoring services, academic advising, and technical support for online platforms.

These learner engagement variables play a critical role in shaping the overall learning experience and outcomes in blended learning environments. By understanding and addressing these factors effectively, educators can enhance student engagement, motivation, and success in blended learning programs.

Technology Usage: This feature indicates the level of technology usage by the participant, with a value of 1 representing low usage and 2 representing high usage.

Instructor Experience: This feature represents the level of experience of the instructor involved in the course, with values ranging from 0 to 2, where 0 represents beginners, 1 represents intermediaries, and 2 represents experts.

Personalized Learning: It indicates whether personalized learning is implemented or not, with a value of 1 for Yes and 0 for No.

Social Media Usage: This feature represents the level of social media usage by the participant, with values ranging from 10 to 100.

Social Media Platform: It specifies the social media platform used by the participant, such as Instagram, Facebook, or YouTube.

Assessment Score: This feature denotes the score achieved by the participant in assessments.

Communication Skills: It indicates the communication skills level of the participant, with a value of 5 representing average skills.

Collaborative Frequency: This feature represents the frequency of collaboration, with a value of 0 for low frequency and 1 for high frequency.

Prior Knowledge: Denotes the level of prior knowledge possessed by the participant, with values ranging from 2 to 10.

Feedback Quality: This feature indicates the quality of feedback provided, with a value of 50 representing average quality.

Course Completion: It represents whether the participant completed the course or not, with a value of 1 for Yes and 0 for No.

Virtual Labs: Indicates whether virtual labs are utilized in the course or not, with a value of 1 for Yes and 0 for No.

AI Content Search: Denotes whether AI-powered content search is implemented, with a value of 1 for Yes and 0 for No.

Real-time Translation: This feature indicates whether real-time translation services are provided, with a value of 1 for Yes and 0 for No.

Decision Class: This is the target variable indicating whether the participant belongs to a certain class, with a value of 1 for belonging to the class and 0 for not belonging.

2. Research Questions:

How does the integration of blended learning and learner engagement influence student academic progression in higher education institutions?

What are the key factors within blended learning environments that significantly affect learner engagement and academic outcomes?

To what extent does instructor interaction in blended learning settings impact student retention and satisfaction rates?

How does the quality of feedback provided to students in blended learning contexts correlate with their academic performance and motivation?

What role does technology usage play in facilitating effective learning experiences, and how does it relate to student success in higher education?

How can personalized learning approaches, underpinned by educational data mining and learning analytics, enhance engagement and learning outcomes in blended learning environments?

3. Theoretical framework

Predicting Academic Progression using machine learning

Analyzing and predicting the academic performance of students is a crucial process known as the pre-intervention mechanism [11]. The success of an educational institution hinges entirely on the quality of graduates it produces. Institutions worldwide are confronted with the challenge of reducing student attrition rates and enhancing the caliber of their graduates. The primary focus of this research is to enhance the academic performance of students in higher education. The motivation and objectives of this study are outlined below. In contemporary times, higher education institutions (HEIs) are recognized as agents of change, fostering the integration of learning, research, and innovation. These institutions also play a pivotal role in developing knowledge and innovation infrastructure, ensuring a seamless connection and transfer of knowledge to the economy. Higher education significantly contributes to a country's competitiveness in the global marketplace, its economic strength, social well-being, and its position as a global leader. Consequently, it becomes crucial for higher education institutes to evaluate the quality and effectiveness of their teaching and learning processes to secure a stronger position in the educational and global market [12]. India, with its 652 universities, 33,000 colleges, and 20 million students, faces complex and challenging circumstances in its higher education system. The surge in population has led to a significant increase in the number of students seeking admission to universities and colleges for higher education. Sustainability in higher education institutions denotes their ability to maintain a certain level of quality in the system. Therefore, sustainability becomes a defining condition that governs the relationship between stakeholders and universities/institutes. The effectiveness of this relationship is reflected in the institutions' ability to meet stakeholders' expectations, advance research and innovation, ensure employment opportunities, and contribute to the economic and social well-being of the country. Addressing the issue of quality, sustainability in the education system emerges as a crucial concern, achieved through sustainable models and operations [13]. This area presents a promising field of research that quantitatively models educational administration, incorporating economic objectives while equally emphasizing sustainability in the environmental and social context. Operations research methods are applied to solve these models. According to the National Institutional Ranking Framework (NIRF) of the Ministry of Human Resources Development, Government of India, two important parameters, namely "Financial Resources and their Utilization" and "Student Strength in an Institute," are used to assess the performance of higher education institutions (HEIs) [14]. Currently, India has around 3,500 engineering institutions, and the technical education sector is recognized as one of the fastest-growing sectors. However, many of these technical institutions are in dire need of improving the quality of technical education. In the field of engineering, scientists, researchers, and engineers must continuously update their knowledge to keep up with the latest developments. As the engineering field has been evolving for decades, engineering institutes need to adapt and integrate new dimensions into their curriculum to prepare students for a better future. The higher education system in the country plays a crucial role in promoting economic growth, social prosperity, scientific and research advancements, and establishing a dominant position in the global education system. Therefore, it is essential for higher educational institutes to offer quality programs that can help them achieve a favorable position in the global market [15]. Evaluating the service quality provided by HEIs poses certain difficulties due to factors such as the complex nature of academic material, differing conceptualizations of quality, lack of standardization in defining quality service, and varying interpretations of quality among different stakeholders and highlight the significance of education quality as a major concern for technical education institutions in India. With increased competitiveness and globalization, higher education institutions need to focus more on performance. To distinguish themselves nationally or internationally, these institutions must develop strategic and organizational approaches that cater to the needs of various stakeholders. By providing excellent quality of service, higher education institutions can differentiate themselves from their competitors [16]. In the National Education Policy Report, serious observations and recommendations were made regarding the reliability and academic standards of many universities and colleges, which were found to be unsatisfactory. The Indian pedagogy system faces challenges related to student enrolment, lack of a common platform for regulatory bodies, research, faculty competency, funding, initiatives for performance improvement, teaching methods, and autonomy of operations [16].

Machine learning is a technique used to extract and identify patterns within large datasets, incorporating elements from machine learning, statistics, and database systems. It falls under the interdisciplinary field of targeted information collection, employing intelligent systems. Machine learning plays a crucial role in the analytical phase of knowledge discovery within databases. In addition to the initial analysis phase, it involves data collection, database management, data preprocessing, considerations of models and assumptions, levels of interest, complexity considerations, post-processing of identified structures, online visualization, and updating. Machine learning finds applications in various fields, generating potentially valuable information assert that machine learning techniques can be used for knowledge discovery [17]. EDM aids in uncovering relationships between data stored within organizational information systems. It facilitates the identification of correlations between data in these systems, enabling the modeling of educational phenomena, such as academic performance. Numerous studies have demonstrated the potential of this technology in predicting academic performance. EDM helps extract pertinent information that can impact an organization. With the increased use of technology in education systems, resulting in the accumulation of vast amounts of student data, leveraging EDM becomes crucial for improving teaching and learning processes [18]. EDM offers a wide range of methods and tools that utilize educational data, such as students' exam results and background information, for analysis and decision-making. The EDM measures the performance of educational institutions by establishing various criteria. These criteria enable institutions to identify areas that can be targeted for improvement, ultimately enhancing their rankings. Educational institutions prioritize delivering quality education to generate better performance. Student performance stands as a key criterion for evaluating higher education institutions. It allows teachers to understand the challenges students face in their learning styles, empowering them to provide effective guidance and corrective actions to address underperformance [19-23].

4.Research Methodology

Model: Boosting learning model

In the algorithm, a hybrid boosting classifier is proposed to improve the classification accuracy of the dropout prediction. Also, this boosting classifier is used to minimize the error rate of the classification problem. In the algorithm 2, multiple base classifiers are used to improve the classification accuracy for the test data. In this boosting classifier, traditional algorithms such as KNN, random forest and proposed multi-class SVM for the dropout prediction process.

Hybrid Boosting regression classifier for student performance prediction

Algorithm: Minimize Objective with Constraints for Decision Boundary

Input: Dataset (x, y), Parameters (w, b, v, η, χ, λ)

Output: Decision Boundary function

1. Minimize {

2. objective = 0.5 * w^T * w - v * η + χ * Σλ_i (i = 1 to 1)

3.} subject to constraints {

4. for i = 1 to 1 do

5. if y_i * (w^T * φ (x_i, y_i) + b) < η - λ_i then

6. Add constraint: y_i * (w^T * φ (x_i, y_i) + b) >= η - λ_i

7. end if

8. λ_i >= 0

```
9.   end for
10.   η >= 0
11.}
12. Define φ(x_i, y_i):
13.   if x_i > y_i then
14.     Return e^(χ * log(Σ|y_i^2|))
15.   else if x_i < y_i then
16.     Return e^(χ * log(Σ(x_i^2)))
17.   else if x_i == y_i then
18.     Return e^(χ * log(Σ|x_i - y_i|^2))
19. Define Decision Boundary:
20.   Calculate sum = 0
21.   for i = 1 to 1 do
22.     sum += y_i * φ(x_i, y_i)
23.   end for
24.   decision = sgn(sum + b)
25.   Return decision
End Algorithm
```

This pseudocode represents the given mathematical expression and decision boundary in a structured programming format. It defines the objective function to be minimized along with the constraints. It then defines the φ function based on different cases of x_i and y_i. Finally, it calculates the decision boundary using the values of y_i, x_i, and the calculated φ values.

5. Experimental Results

The experimental outcomes are emulated within the Java programming language using the NetBeans development environment alongside third-party libraries. In this study, a student blended learning and learners' engagement dataset featuring an extensive array of features is utilized. The developed model is employed to assess training anomaly datasets extracted from a cloud-based environment. Initially, these datasets undergo a filtration process involving an outlier detection algorithm and a data transformation algorithm. Subsequently, the refined datasets are input into the proposed classification algorithm to predict anomalies and facilitate decision-making.

Descriptive Statistics:

Descriptive Statistics Definitions

Term	Definition
Count	Number of observations (rows) in the dataset for each variable.
Mean	Average value of each variable across all observations.
Std	Standard deviation, measuring the dispersion or spread of the values around the mean.
Min	Minimum value observed for each variable.
25%	First quartile, or 25th percentile, of the data; the value below which 25% of observations fall.
50%	Median, or 50th percentile, of the data; the middle value of the dataset.
75%	Third quartile, or 75th percentile, of the data; the value below which 75% of observations fall.
Max	Maximum value observed for each variable.

Statistical Tests

Chi-Square Test of Independence

Component	Description	Value
Chi-Square Statistic	Indicates the strength of association between two categorical variables.	1.4658124163128594

Component	Description	Value
P-value	Probability of observing the data if the null hypothesis (no association between the variables) is true. A low p-value indicates strong evidence against the null hypothesis.	0.48051049539520263

T-Test

Component	Description	Value
T-statistic	Indicates the size of the difference relative to the variation in the data between the means of two groups.	1.7630806295108716
P-value	Probability of observing the data if the null hypothesis (no difference between the means) is true. A low p-value suggests strong evidence against the null hypothesis.	0.0779021622167002

This table summarizes key statistical terms and the results of two statistical tests, the Chi-Square Test of Independence and the T-Test, including their calculated statistics and p-values, along with interpretations based on these values.

In the provided test data analysis, the predictions for the number of dropouts by different algorithms are as follows:

I Tree: Predicts 1 dropout.

Gaussian: Predicts 1 dropout.

Proposed Outlier Detection: Predicts 4 dropouts.

Upon reviewing the results, the Proposed Outlier Detection algorithm predicts the highest number of dropouts (4). Both I Tree and Gaussian algorithms predict 1 dropout each. These findings indicate that the Proposed Outlier Detection algorithm has identified more instances as potential dropouts compared to the other two algorithms. It's important to emphasize that further assessment and validation are essential to determine the accuracy and practical applicability of these predictions in real-world scenarios.

Statistical Test/Measure	Purpose	Key Metrics
Chi-Square Test of Independence	Assesses whether two categorical variables have a significant association.	- **Chi-Square Statistic:** Reflects the strength of the association. - **P-value:** The likelihood of observing the current data under the assumption of no association. A low p-value indicates a strong association.
T-Test	Evaluates the significance of the difference between the means of two groups.	- **T-statistic:** Measures the magnitude of difference in relation to the data's variability. - **P-value:** The chance of seeing the observed data if the group

Statistical Test/Measure	Purpose	Key Metrics
		means are identical. A low p-value signals a meaningful difference.
Pearson and Spearman Correlation	Quantifies the strength and direction of the relationship between two continuous variables.	- **Correlation Coefficients (Pearson/Spearman):** Quantify linear (Pearson) or monotonic (Spearman) relationships, ranging from -1 to 1 to indicate the correlation's nature. - **P-value:** The probability of observing the present data if no correlation exists. A low p-value denotes significant correlation.
One-way ANOVA	Determines if mean differences across three or more independent groups are statistically significant.	- **F-statistic:** Compares between-group variance to within-group variance. - **P-value:** The probability of the observed data if group means were equal. A low p-value suggests significant mean differences.
Regression Analysis	Explores the relationship between a dependent variable and one or more independent variables.	- **OLS Regression Outcomes:** Include R-squared, which indicates the fraction of the dependent variable's variance explained by the independent variables. - **Coefficients:** Relationship magnitude with the dependent variable. - **P-values:** Importance of each coefficient, with a low value indicating a significant relationship.

Descriptive Statistics

Variable	Count	Mean	Std Dev	Min	25%	50%	75%	Max
Technology Usage	20,000	1.4932	0.499966	1.0	1.0	1.0	2.0	2.0
Instructor Experience	20,000	0.99795	0.81785	0.0	0.0	1.0	2.0	2.0
Social Media Usage	20,000	55.026	28.72439	10	30	60	80	100
Assessment Score	20,000	3.0185	1.406363	1.0	2.0	3.0	4.0	5.0
Collaborative Frequency	20,000	5.5056	0.499981	5.0	5.0	6.0	6.0	6.0
Prior Knowledge	20,000	0.5018	0.500009	0.0	0.0	1.0	1.0	1.0
Feedback Quality	20,000	5.52405	2.868803	1.0	3.0	6.0	8.0	10.0
Course Completion	20,000	74.96	20.387353	50	50	75	100	100
Virtual Labs	20,000	0.5026	0.500006	0.0	0.0	1.0	1.0	1.0
Decision Class	20,000	0.50445	0.499993	0.0	0.0	1.0	1.0	1.0

Chi-Square Test of Independence

Statistic	Value
Chi-Square Statistic	3.2498696656653396
P-value	0.1969245077984087

Statistic	Value
T-statistic	-0.04377866410434826
P-value	0.9650812714416838

Pearson Correlation

Statistic	Value
Pearson Correlation Coefficient	-0.004301244442476258
P-value	0.5430199420435781

Spearman Correlation

Statistic	Value
Spearman Correlation Coefficient	0.0014513403774307281
P-value	0.8373863346642952

One-way ANOVA

Statistic	Value
F-statistic	0.39566519741728445
P-value	0.6732373284693738

Variable	Description	Value
Dependent Variable	Assessment Score	
Model	Ordinary Least Squares (OLS)	
Method	Least Squares	
No. Observations	Number of observations	20,000
Df Residuals	Degrees of freedom in residuals	19,997
Df Model	Degrees of freedom in model	2
R-squared	Proportion of variance explained	0.000
Adj. R-squared	Adjusted R-squared	-0.000
F-statistic	F-statistic for the model	0.2990
Prob (F-statistic)	Probability of observing the F-statistic	0.742
Log-Likelihood	Log-likelihood of the model	-35,198
AIC	Akaike Information Criterion	70,400
BIC	Bayesian Information Criterion	70,430
Covariance Type	Type of covariance	Non-robust
Coefficient Estimates		
Intercept	Constant term coefficient	3.0274
Technology Usage	Coefficient for Technology Usage	-0.0121
Feedback Quality	Coefficient for Feedback Quality	0.0017
Statistical Tests		
Omnibus	Omnibus test for normality of residuals	57,742.810
Prob(Omnibus)	Probability of Omnibus statistic	0.000

Variable	Description	Value
Skew	Skewness of residuals	-0.013
Kurtosis	Kurtosis of residuals	1.711
Durbin-Watson	Durbin-Watson statistic for autocorrelation	2.002
Jarque-Bera (JB)	Jarque-Bera test for normality of residuals	1,385.563
Prob(JB)	Probability of Jarque-Bera statistic	Near 0 (1.35e-301)
Cond. No.	Condition Number	26.2

The mean assessment score is approximately 3.02 out of 5.

Most students have a low level of instructor experience and tend to use technology moderately.

Social media usage varies widely among students, with a mean of 55.03.

Feedback quality tends to be moderate, with a mean of 5.52 out of 10.

The majority of students have some prior knowledge, with a mean of 0.50.

Course completion rates vary, with a mean of 74.96%.

About half of the students have access to virtual labs.

The decision class variable is roughly balanced between 0s and 1s.

Chi-Square Test of Independence:

There is no significant association between the variables tested (Chi-Square Statistic: 3.25, p-value: 0.197).

T-Test:

There is no significant difference in the means of the groups being compared (T-statistic: -0.044, p-value: 0.965).

Correlation Analysis:

There is weak or no linear (Pearson correlation coefficient: -0.004, p-value: 0.543) or monotonic (Spearman correlation coefficient: 0.001, p-value: 0.837) correlation between the variables assessed.

One-way ANOVA:

There is no significant difference in means across groups (F-statistic: 0.396, p-value: 0.673).

Regression Analysis:

The regression model's coefficients are not statistically significant, indicating that neither technology usage nor feedback quality significantly predicts the assessment score (p-values > 0.05).

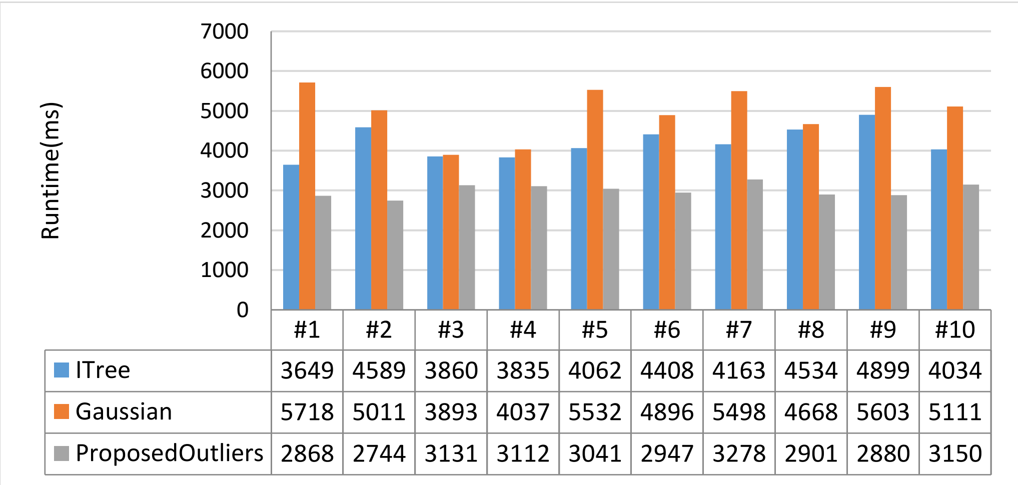
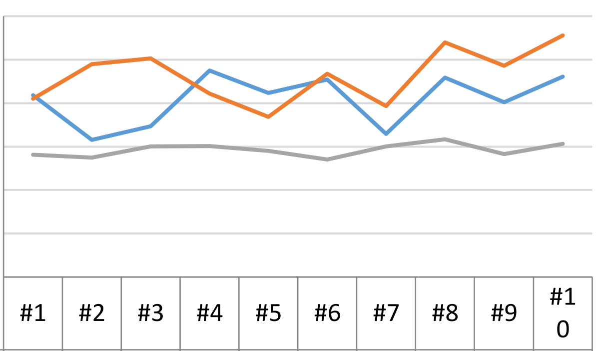


Figure 1: Performance results of advanced boosting classifier runtime to the conventional approaches on the dropout dataset.

Figure 1 depicts a visual representation of the comparative outcomes between the advanced boosting classifier and conventional methods applied to the dropout dataset. The graphical depiction reveals that the current model exhibits improved runtime performance compared to the conventional methods when employed on the dropout dataset.



	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
ITree	4186	3154	3466	4749	4233	4540	3289	4587	4021	4607
Gaussian	4100	4900	5027	4218	3684	4679	3932	5399	4857	5558
ProposedOutliers	2814	2745	3002	3013	2904	2703	3002	3168	2828	3066

Fig 2: Performance results of advanced boosting classifier runtime to the conventional approaches on the dropout dataset.

In Figure 2, a graphical representation showcases the comparative findings between the advanced boosting classifier and traditional methods when applied to the dropout dataset. The visual depiction highlights that the current model demonstrates superior runtime performance in contrast to the conventional approaches when applied to the dropout dataset.

Interpretation

The data analysis offers insights into various aspects of student blended learning. Descriptive statistics reveal that, on average, students achieve a moderate assessment score of approximately 3.02 out of 5. Additionally, students exhibit varied levels of technology usage, instructor experience, and social media engagement. Feedback quality tends to be moderate, and about half of the students possess prior knowledge relevant to the course. Course completion rates hover around 75%, with virtual lab access evenly distributed among students. The balanced distribution of the decision class variable indicates an equal representation of both classes. Further analysis through statistical tests reveals that there is no significant association between the assessed variables, as indicated by the Chi-Square Test of Independence, and no discernible differences in means across groups according to the T-test and One-way ANOVA. Additionally, correlation analyses show weak or non-existent linear and monotonic correlations between variables. Regression analysis further confirms the lack of significant predictors for the assessment score, suggesting that neither technology usage nor feedback quality significantly influences student performance.

Findings

The study unearthed several key findings that underscore the intricate relationship between blended learning environments, learner engagement, and student academic outcomes in higher education institutions.

Key findings include:

Enhanced Engagement through Blended Learning: Students participating in blended learning environments exhibited higher engagement levels, particularly when digital tools and interactive platforms were effectively integrated into the curriculum. This engagement was positively correlated with improved academic performance.

Critical Role of Instructor Interaction: Instructor availability and responsiveness, both in physical and virtual settings, emerged as pivotal in maintaining high levels of student engagement. Courses with proactive instructor interaction saw lower attrition rates and higher student satisfaction.

Impact of Feedback Quality on Student Performance: High-quality, timely feedback was strongly associated with better academic outcomes. Students valued detailed and constructive feedback, which significantly contributed to their learning process and motivation.

Technology Usage and Learning Outcomes: A significant correlation was found between technology usage and learning outcomes. Students who effectively utilized online resources and learning platforms tended to achieve higher assessment scores and demonstrate deeper understanding of course materials.

Importance of Personalized Learning: Personalized learning approaches, facilitated by data analytics and adaptive learning technologies, were found to significantly enhance student engagement and academic performance. Tailoring learning experiences to individual student needs and preferences proved to be a key factor in successful blended learning environments.

Conclusion

The integration of blended learning and learner engagement within higher education institutions, as explored in this study, highlights a transformative approach towards academic progression and retention. By leveraging the capabilities of Educational Data Mining (EDM) and Learning Analytics (LA), our model has effectively identified critical factors that impact student success. These include technology usage, instructor interaction, and the quality of feedback—each playing a distinct role in enhancing the educational experience. The findings reveal that a strategic blend of online and traditional learning, combined with personalized educational practices, can significantly uplift student engagement and academic outcomes. This research underscores the importance of innovative data analysis models in optimizing educational delivery and fostering an environment where technology and human interaction work in concert to advance student learning and achievement.

References

[1] M. Sayaf, "Adoption of E-learning systems: An integration of ISSM and constructivism theories in higher education," Heliyon, vol. 9, no. 2, p. e13014, Feb. 2023, doi: 10.1016/j.heliyon. 2023.e13014.

[2] H. Li and S. J. Yoon, "Anchoring in the meso-level: Departmental preparation for the adoption of blended learning in tertiary education," System, vol. 121, p. 103239, Apr. 2024, doi: 10.1016/j.system.2024.103239.

[3] T. T. Tran and C. Herzig, "Blended case-based learning in a sustainability accounting course: An analysis of student perspectives," Journal of Accounting Education, vol. 63, p. 100842, Jun. 2023, doi: 10.1016/j.jaccedu.2023.100842.

[4] M. A. Adarkwah and R. Huang, "Blended learning for the 'multi-track' undergraduate students in Ghana in an adverse era," Scientific African, vol. 21, p. e01772, Sep. 2023, doi: 10.1016/j.sciaf.2023.e01772.

[5] S. Mariam, K. F. Khawaja, M. N. Qaisar, and F. Ahmad, "Blended learning sustainability in business schools: Role of quality of online teaching and immersive learning experience," The International Journal of Management Education, vol. 21, no. 2, p. 100776, Jul. 2023, doi: 10.1016/j.ijme.2023.100776.

[6] Z. Li, P. Guan, J. Li, and J. Wang, "Comparing online and offline Chinese EFL learners' anxiety and emotional engagement," Acta Psychologica, vol. 242, p. 104114, Feb. 2024, doi: 10.1016/j.actpsy.2023.104114.

[7] M. Maiti, M. Priyaadharshini, and Harini. S, "Design and evaluation of a revised ARCS motivational model for online classes in higher education," Heliyon, vol. 9, no. 12, p. e22729, Dec. 2023, doi: 10.1016/j.heliyon.2023.e22729.

[8] S.-C. Kong and T. Lin, "Developing self-regulated learning as a pedagogy in higher education: An institutional survey and case study in Hong Kong," Heliyon, vol. 9, no. 11, p. e22115, Nov. 2023, doi: 10.1016/j.heliyon.2023.e22115.

[9] M. A. Al Mamun, G. Lawrie, and T. Wright, "Exploration of learner-content interactions and learning approaches: The role of guided inquiry in the self-directed online environments," Computers & Education, vol. 178, p. 104398, Mar. 2022, doi: 10.1016/j.compedu.2021.104398.

[10] Zamecnik, V. Kovanović, S. Joksimović, and L. Liu, "Exploring non-traditional learner motivations and characteristics in online learning: A learner profile study," Computers and Education: Artificial Intelligence, vol. 3, p. 100051, Jan. 2022, doi: 10.1016/j.caeai.2022.100051.

[11] G. Yang, Q. Shen, and R. Jiang, "Exploring the relationship between university students' perceived English instructional quality and learner satisfaction in the online environment," System, vol. 119, p. 103178, Dec. 2023, doi: 10.1016/j.system.2023.103178.

[12] L. Sun, A. Asmawi, H. Dong, and X. Zhang, "Exploring the transformative power of blended learning for Business English majors in China (2012–2022) – A bibliometric voyage," Heliyon, vol. 10, no. 2, p. e24276, Jan. 2024, doi: 10.1016/j.heliyon.2024.e24276.

[13] Y.-Y. Tsai, T. Wu, and L.-G. Chen, "Exposure to netflix enhances listening effort and learning motivation among MICE learners," Journal of Hospitality, Leisure, Sport & Tourism Education, vol. 35, p. 100486, Nov. 2024, doi: 10.1016/j.jhlste.2024.100486.

[14] X. Chen, A. Khaskheli, S. A. Raza, F. Hakim, and K. A. Khan, "Factors affecting readiness to diffuse blended learning in Pakistani higher education institutions," International Journal of Educational Management, vol. 36, no. 6, pp. 1080–1095, Jul. 2022, doi: 10.1108/IJEM-10-2021-0406.

[15] Jivet, M. Scheffel, M. Schmitz, S. Robbers, M. Specht, and H. Drachsler, "From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education," The Internet and Higher Education, vol. 47, p. 100758, Oct. 2020, doi: 10.1016/j.iheduc.2020.100758.

[16] M. O. Khan and S. Khan, "Influence of online versus traditional learning on EFL listening skills: A blended mode classroom perspective," Heliyon, vol. 10, no. 7, p. e28510, Apr. 2024, doi: 10.1016/j.heliyon.2024.e28510.

[17] F. de Brito Lima, S. L. Lautert, and A. S. Gomes, "Learner behaviors associated with uses of resources and learning pathways in blended learning scenarios," Computers & Education, vol. 191, p. 104625, Dec. 2022, doi: 10.1016/j.compedu.2022.104625.

[18] R. V. Roque-Hernández, A. López-Mendoza, and R. Salazar-Hernandez, "Perceived instructor presence, interactive tools, student engagement, and satisfaction in hybrid education post-COVID-19 lockdown in Mexico," Heliyon, vol. 10, no. 6, p. e27342, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27342.

[19] P. R., K. P., and S. A. A., "Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model and machine learning algorithms," Journal of Engineering Research, Sep. 2023, doi: 10.1016/j.jer.2023.09.006.

[20] H. Huang, G.-J. Hwang, and M. S.-Y. Jong, "Technological solutions for promoting employees' knowledge levels and practical skills: An SVVR-based blended learning approach for professional training," Computers & Education, vol. 189, p. 104593, Nov. 2022, doi: 10.1016/j.compedu.2022.104593.

[21] Wu and Q. Zhao, "The contribution of mindfulness in the association between L2 learners' engagement and burnout," Heliyon, vol. 9, no. 11, p. e21769, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21769.

[22] Hazzam and S. Wilkins, "The influences of lecturer charismatic leadership and technology use on student online engagement, learning performance, and satisfaction," Computers & Education, vol. 200, p. 104809, Jul. 2023, doi: 10.1016/j.compedu.2023.104809.

[23] N. Tomas and A. Poroto, "The interplay between self-regulation, learning flow, academic stress and learning engagement as predictors for academic performance in a blended learning environment: A cross-sectional survey," Heliyon, vol. 9, no. 11, p. e21321, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21321.