International Journal of Evaluation and Research in Education (IJERE)
Vol. 14, No. 5, October 2025, pp. 3872~3881
ISSN: 2252-8822, DOI: 10.11591/ijere.v14i5.33815

Predicting student performance and identifying learning behaviors using decision trees and K-means clustering

Md. Mahadhi Hasan[1], Md Nakibul Islam[2], Md Ikramul Haque Nirjon[2], Md Sharif Uddin[2], Md. Muntasir Mamun[1], Zaheed Alam Munna[3], Al Mahmud Rumman[1]
[1]Department of English and Humanities, University of Liberal Arts Bangladesh, Dhaka, Bangladesh
[2]School of Information Technology, Washington University of Science and Technology, Alexandria, United States
[3]Department of English, Southeast University, Dhaka, Bangladesh

Article Info

Article history:

Received Nov 27, 2024
Revised Apr 10, 2025
Accepted May 9, 2025

Keywords:

Bangladesh
Decision trees
Higher education
K-means clustering
Learning analytics
Learning behaviors
Student performance

ABSTRACT

The insufficiency of a strong mechanism to measure student performance and learning behavior has been pointed out as a result of the expansion of higher education in Bangladesh. The objectives of the study are to predict students' performance and recognize unique learning behaviors in the Bangladeshi higher education contexts by applying decision trees and K-means clustering methods. Validity and reliability of the results are ensured by following methods: 10-fold cross-validation for the decision tree model and Silhouette score assessment for the K-means clustering model, thus improving the predictive accuracy and differentiation of clusters. The study is based on a dataset of student records numbering 1,200, researching factors such as attendance (91.22%), exam results (mean 83.54%), completed assignments (mean 80.54%), and age (mean 23.47). Learning analytics theory is used since it is crucial to apply data to enhance the understanding and effectiveness of learning processes. The decision tree model showed excellent performance with high rates in precision, recall, and F1-scores, which were all at 0.99 for the evaluated performance measures, hence increasing its good predictive power. K-means clustering analysis grouped the students into three distinct groups: active learners, passive learners, and at-risk students. This research urges the adaptation of data mining methodologies within the framework of higher education and strongly emphasizes the important role that an early identification of at-risk students can play. This research is a contribution to the learning analytics area, and it further proves the applicability of data mining methods in predicting academic performance and improving education outcomes in developing contexts.

Corresponding Author:

Md. Mahadhi Hasan
Department of English and Humanities, University of Liberal Arts Bangladesh
688 Beribadh Road, 1207 Dhaka, Bangladesh
Email: mahadhi4bd@gmail.com

1. INTRODUCTION

The rapid advancement of higher education in Bangladesh has given increased impetus to sophisticated techniques for predicting student performance and understanding learning behavior [1]–[4]. Traditional assessment methods in education have been widely criticized as insufficient to capture the complexity of student learning outcomes, especially in light of growing diversity among higher education students and increasing integration of technology into higher education [1]–[4]. The purpose of this research is to apply data mining techniques, specifically decision trees and K-means clustering, for predicting students' performance as well as to identify unique learning patterns in the context of the higher education system of Bangladesh. The methodology to be used has its basis on the principles of learning analytics theory, which emphasizes the use of data to understand and enhance learning systems. This is achieved through a focus on collecting, analyzing, and interpreting learners' achievement, engagement, and behavioral data to make informed decisions based on the data gathered [5]. These methodologies are increasingly being implemented in educational research because of their ability to handle large databases and uncover hidden patterns, therefore greatly helping to shed light on the factors impacting academic achievement [6], [7].

Previous literature has consistently found strong evidence of the influence of a number of academic and behavioral factors—like attendance, test scores, and submission of assignments—on student outcomes. For instance, a study established that class participation acts as a mediating variable between attendance and academic success, hence drawing particular interest to the fact that engagement is one of the leading performance predictors [8]. In line with this, another study found that test scores represent a strong predictor of academic success, especially when used in combination with past academic records [9]. All too often, however, these factors are examined in isolation, leaving a gap in our understanding of how they interact and the combined effects on student outcomes. Other studies did not explore Bangladeshi education context in terms of student performance and learning behavior. This gap is particularly important in Bangladesh, where the introduction of English-medium instruction has been blended with digital learning technologies in changing the face of education [10], [11]. While the presence of English as a medium of instruction (EMI) in the higher education context has often been associated with increased opportunities and better employment prospects, it also poses challenges of academic engagement and the effective integration of digital tools [10], [12].

Scholars showed that class participation played a mediating role in the relationship between attendance and academic performance, thereby underscoring the very important role of engagement in predicting educational outcomes [8]. Another study found that a supportive school climate reduces truancy, which might be of relevance in the Bangladeshi setting [13]. In addition, another study confirmed that attendance positively impacts academic performance, particularly for students with lower achievement levels [14]. Scholars hypothesized that university attendance, while often associated with better critical and creative thinking, resulted in no such association from their study results, thereby suggesting the need for further investigation [15]. Taken together, these results provide support for the data mining applications of techniques such as decision trees and K-means clustering in the analysis of academic engagement and prediction of student success. In a related study, it is demonstrated the effectiveness of machine learning methods, using midterm examination scores as a principal variable to effectively predict final examination grades [6]. Another study emphasized student engagement, where students' scores are often conditioned on things like attending classes and using resources, thus making them reflective of bigger academic behaviors [16]. Scholars also noted that resilience and grit, though important, had smaller correlations with academic outcomes than academic engagement regarding exam performance [7].

These research findings support the fact that examination scores are one of the most critical variables in predicting student performance, which aligns with the goal of the current study to predict academic success and identify learning patterns within the higher education setting in Bangladesh. Research places much emphasis on the submission of assignments with regard to academic success and as one of the major indicators of student performance. A scholar proved that in-class group assignments (IGAs) improved the performance of underachieving students in accounting courses, particularly on examination questions related to those assignments, which may mean that assignments can serve students as a useful mechanism for strengthening their learning [17]. A related study observed how students use the flexibility of assignment submission times and showed that delayed submissions, regardless of any time limits, negatively influence academic performance [18].

Studies have analyzed the interaction of age with other factors, such as motivation and personality, which collectively determine learning strategies and outcomes [19]–[23]. Scholars also found that age determines students' engagement in peer interactions and suggested that younger learners benefit more in interactive learning contexts [14]. The results indicate that age serves as a significant factor affecting academic behaviors, which may play a vital role in forecasting student performance and recognizing unique learning patterns. The conventional teaching methods remain prevalent, with a significant number of educators lacking familiarity with andragogical or heutagogical frameworks, thereby constraining student-centered educational practices [12]. Scholars also noted that the COVID-19 pandemic accelerated the adoption of digital learning [11]; however, students' tendency to engage with technology is a vital factor determining its effectiveness [20]–[23].

Besides academic sources, age has also been established as a determinant variable that influences learning behaviors. Studies indicate that young learners tend to show different trends in regard to their participation in activities and academic performance compared to older learners, which essentially suggests that age is a critical factor in shaping learning behaviors [13], [14]. Not much research, however, exists to establish the interaction between age and these other variables: attendance, examination scores, and assignment submission rates, in the context of higher education in Bangladesh. The present study tries to fill this gap by encompassing these factors in a comprehensive model of academic achievement while unveiling fundamental trends in learning. Even though digital resources create new learning opportunities, the extent to which students are willing and able to engage with these platforms varies significantly and this heterogeneity should be considered in predicting student success [11].

This study uses decision trees in identifying, in a clear and distinguishable manner, the most important student achievement predictors, hence allowing a much deeper understanding of the factors that lead to success in higher education in Bangladesh. K-means clustering is applied to identify different learning behavior patterns by students, allowing grouping of students in respective classes based on their level of academic engagement. This methodology is in line with the findings reported in past studies, which argue that student engagement, in terms of attendance and participation, plays a substantial role in determining academic performance but remains under-explored in data mining methodologies [15], [16].

This study could be an important addition to the literature, as it fills the current gap in providing a broad, data-driven assessment of the major factors determining student achievement and learning behaviors in Bangladesh. Such integration helps to make the research much clearer in the elucidation of the interrelations among attendance, examination scores, and assignment submission rates, age, and digital engagement within a singular model of explaining academic performance. This study answers the two research questions:

i) How can decision trees be used to predict student performance in Bangladeshi higher education?
ii) How can K-means clustering identify unique learning behaviors among students in Bangladesh's higher education sector?

2. METHOD

This research uses decision trees and K-means clustering techniques to predict student performance and to identify distinctive learning patterns in a private university in Bangladesh. The study focuses on using student data to answer critical research questions regarding academic outcomes and the classification of learning behavior in Bangladeshi context.

2.1. Data collection and preparation

A sample of 1,200 anonymous student records was used for this present study, covering two academic years from a private university in Bangladesh. The chosen sample size satisfies the suggestions of a scholar, which state that the datasets of 1,000–1,500 records ensure robust predictions for data mining techniques including decision trees and K-means clustering [24]. The main academic and behavioral variables available in this dataset include: age, attendance (%), exam scores, assignment submission (%), and cumulative grade point average (CGPA). Missing values were imputed using mean imputation, following the recommendation of the scholars to avoid bias while preserving data quality [25]. The continuous features such as attendance (%), exam scores, and CGPA were normalized using the min-max scaling to ensure that the features are comparable between each other [26]. The dataset was then split into a training set (80%) and a test set (20%) for model performance evaluation and generalizability.

The inclusion criteria for this study were anonymized student records from a private university in Bangladesh for two academic years. A total of 1,200 records were selected to meet the recommended dataset size of 1,000 to 1,500, thus allowing for accurate predictions for data mining techniques, such as decision trees and K-means clustering. Data entries falling outside the specified academic years or with incomplete information, which might affect the accuracy of the analysis to a great extent, were excluded from the study.

2.2. Decision tree model for predicting student performance

To answer the first research question, the classification and regression trees (CART) algorithm was applied [27]. The dependent variable, performance category, was derived from CGPA and categorized into three classes: high performance (CGPA>3.75), average performance (CGPA 2.5–3.75), and low performance (CGPA<2.5). The predictor variables included age, attendance (%), exam scores and assignment submission (%), thus covering both academic and behavioral dynamics. The analysis supported that high levels of attendance and exam scores are both strong predictors of good performance, which is consistent with previous findings [26]. To avoid overfitting, pruning methods were applied, and model validation was completed based on 10-fold cross-validation. Performance metrics, such as accuracy, precision, recall, and F1 score, showed excellent predictive accuracy, with all the metrics having values over 0.98.

2.3. K-means clustering for identifying learning behaviors

The second research question was analyzed through the use of K-means clustering. The following input variables, attendance (%), exam scores, age, and assignment submission (%) were chosen to represent student discipline and engagement. Through the Elbow method, it was identified that three clusters formed the optimal number [28], a finding further justified by a Silhouette score of 0.54, indicating moderate clustering effectiveness [29]. Data normalization and the removal of outliers were critical to ensuring the spherical assumption of K-means. The clustering revealed three groups: active learners (high attendance and consistent submissions), passive learners (moderate engagement), and at-risk students (low attendance and submissions, despite high exam scores). Validation by CGPA confirmed that the clusters were aligned with academic performance and provided actionable insights into varied learning behaviors.

2.4. Validation and benchmarking of models

To address the issues of reliability and generalizability, models in this study have gone through the validation processes. The decision tree model was validated using 10-fold cross-validation, using such metrics as accuracy, precision, recall, and F1 score to avoid the risk of overfitting and maximize the predictive power of the model [30], [31]. The K-means clustering model was validated by Silhouette score, which provided the distinctions and proximities among all the clusters. Comparing the results from both models to simpler approaches—linear regression, concerning forecasting performance, and manual classification, with regard to the recognition of learning behaviors—showed the following: these analyses, therefore, brought out the increased precision and pattern recognition of the data mining models, hence proving valuable toward the derivation of meaningful information from the data set.

3. RESULTS AND DISCUSSION

3.1. Results

The present research work applies decision trees and K-means clustering to analyze student performance and learning patterns in a university setup. With this, the most important academic indicators and unique learner profiles are given. These are essential for sketching engagement trends in view and enabling strategies toward better educational outcomes in Bangladeshi universities.

3.1.1. Decision tree model results

The decision tree model shows excellent predictive accuracy in the classification of student performance, such as that indicated by precision, recall, and F1-scores, which range between 0.98 and 1.00 for all the categories. Such a measure brings to light the model's ability to lower both Type I and Type II errors while ensuring considerable classification reliability. The support values, indicative of the actual instance distribution in the dataset, further reinforce the model's robustness, which underwent evaluation with a total of 360 cases. The weighted averages for precision, recall, and F1-score all align at 0.99, indicating homogeneous performance across the different students classes. These results underscore the effectiveness of this model in identifying key performance trends; as such, it offers actionable insights for timely intervention and tailored academic assistance. The methodological rigor of the study, incorporating cross-validation and pruning techniques, mitigates the risk of overfitting and enhances the model's generalizability in the context of higher education in Bangladesh. The findings are mentioned in Table 1.

Table 1. Decision tree: student performance metrics

Precision	Recall	F1-score	Support
0.98	0.98	0.98	109.00
0.99	0.99	0.99	180.00
1.00	1.00	1.00	71.00
0.99	0.99	0.99	0.99
0.99	0.99	0.99	360.00
0.99	0.99	0.99	360.00

3.1.2. K-means clustering results

a. Cluster characteristics by academic engagement and performance

K-means clustering analysis pointed out three clear student profiles based on their academic engagement and performance. Cluster 1, or "active learners," consists of younger students (average age 20.72) who have the highest percentage of attendance (91.22%), submitting assignments (80.54%), and high scores on exams (83.54), resulting in a strong average CGPA of 3.15. Cluster 2, "passive learners," has students with moderate engagement in terms of average attendance at 76.63%, average exam scores of 63.78, and assignment submission rates of 64.85% that resulted in a lower CGPA compared to the three clusters, at 2.58. On the other hand, Cluster 3, “at-risk students,” consists of older students (mean age 24.93) who have the lowest attendance (59.62%) and assignment submissions (52.55%), but report very high exam scores (94.34) and relatively moderate CGPA (2.69). These findings, as shown in Table 2, emphasize different learning behaviors, which allow tailored interventions to improve academic success.

Table 2. K-means clustering: academic and performance characteristics

Age	Attendance (%)	Exam scores	Assignment submission (%)	CGPA (1.0–4.0)
20.72	91.22	83.54	80.54	3.15
23.47	76.63	63.78	64.85	2.58
24.93	59.62	94.34	52.55	2.69

b. Within-cluster sum of squares values for k-means clustering

The following are the within-cluster sum of squares (WCSS) values that show the variances within clusters for a different number of clusters, which helps in determining the optimal number of clusters that the K-means algorithm is able to achieve. When the number of clusters increases from 1 to 5, the WCSS values fall drastically, indicating more definable clusters. For one cluster, the WCSS is 4800.00, which reflects high internal variance. With two clusters, WCSS drops to 2698.08; with three, it decreases to 1718.18; and with four, it further decreases to 1146.35. By the fifth cluster, one has a WCSS of 890.02, indicating diminishing returns in the reduction of variance as clusters increase. The steep decline between 1 to 3 clusters suggests that three clusters may balance simplicity with model fit, aligning with the “Elbow” in the Elbow method. This confirms the appropriateness of three clusters for identifying distinct student groups in this study. The findings are presented in Table 3.

Table 3. WCSS for optimal clusters in K-means

Number of clusters	WCSS
1	4800.00
2	2698.08
3	1718.18
4	1146.35
5	890.02

c. Silhouette score for K-means clustering

The Silhouette score obtained from the K-means clustering analysis is 0.54, which indicates moderate clustering quality. It measures the appropriateness of the assignment of data points to their clusters; scores are closest to 1 when the clusters are well defined and compact, while scores near 0 signify overlapping clusters. A score of 0.54, therefore, signals clusters that are well separated but not perfectly so, with some overlap. The achieved cohesion and separation provide adequate reliability for the identification of critical patterns of student learning behaviors, which support further clustering method improvement or additional feature selections.

3.2. Discussion

3.2.1. Performance and learning behavior of Bangladeshi students

The decision tree model results demonstrate perfect predictive accuracy in student performance classification, with precision, recall, and F1-scores all exceeding 0.98 in each class. The model, showed a good capability of correctly classifying students into three classes: high, average, and low performers, based on the variables of attendance, examination scores, and assignment submission. These results align with previous findings, which have identified in-class participation, a feature closely related to attendance, as a significant indicator of academic success [8]. Similarly, scholars have found exam scores to be significantly crucial in measuring student achievement, which aligns with study's findings that show exam performance to be one of the critical features in the performance of the decision tree model [9]. Similarly, the high values of precision and recall from the model give evidence that it equally reduces Type I and Type II errors, which is vital for the consistent classification of student performance. In this way, while the model inherits the principles of learning analytics theory, it does so through data-informed insights, providing early intervention and guidance for informed decision-making in a course of action to improve student outcomes [5].

The model performs exceptionally well, with weighted averages of 0.99 across all measures, underscoring its reliability across a variety of student populations and academic behaviors. This supports the previous work that machine learning algorithms, such as decision trees, can be effectively used in making predictions about academic success based on critical predictors, including examination outcomes [6]. It also employs the most rigorous methods of validation used in this study, such as cross-validation and pruning, to mitigate overfitting and enhance the generalizability of the results. This methodological rigor aligns with scholarly statements on the importance of accurate and stable predictions in the domain of learning analytics [5]. The current paper reports findings that show decision trees, with high classification accuracy, can be an effective tool for predicting students' performance in Bangladeshi higher education, especially in scenarios where traditional assessment methods may poorly represent the complex interplay of factors influencing academic success.

K-means clustering analysis identified three distinct student profiles based on academic engagement and performance, thereby unraveling the multifaceted nature of student behaviors and the determinants leading to academic success. Cluster 1, referred to as “active learners,” comprises younger students with high attendance rates (91.22%), high submission rates for assignments (80.54%), and high examination scores (83.54%), resulting in a strong average CGPA of 3.15. These observations in this cluster are consistent with the findings of preceding scholars, who concluded that high levels of engagement, particularly in terms of attendance and participation, are strongly correlated with academic success [8]. Based on learning analytics theory, the identified patterns confirm that engagement metrics help predict academic outcomes, thereby providing actionable insights for early interventions [5]. In sharp contrast, Cluster 2, labeled “passive learners,” comprises students with moderate levels of engagement, characterized by a 76.63% attendance rate, an average exam score of 63.78, and a 64.85% rate of assignment submissions. Correspondingly, the lower average cumulative GPA of 2.58 is matched by this trend. This result is consistent with the broader trend in educational scholarship, which suggests that lower levels of engagement result in lower levels of learning [15]. These findings emphasize the importance of using data analytics for managing student engagement, as it is found that students exhibiting disengagement tend to have lower academic achievements. This calls for tailor-made strategies to bring these students back into the learning process. Lastly, Cluster 3, “at-risk students,” comprises the oldest students with the lowest attendance rates (59.62%) and submission rates of assignments (52.55%), yet surprisingly high exam scores (94.34%), resulting in a moderate cumulative GPA of 2.69.

Some students, although not exhibiting much engagement in other aspects, may have achieved higher results in exams by using cramming techniques or other focused study habits [7]. This could highlight the inherent complexity associated with student performance and demonstrate the need to apply learning analytics theory to scrutinize these behaviors [5]. This creates different implications for students who have low levels of engagement, as they may need various kinds of support, such as methods for test preparation or possibly even training in time management. This demonstrates the importance of defining different learning behaviors when intervening to improve outcomes for students. One of the key goals of learning analytics is to inform decision-making and refine educational practices.

Furthermore, the WCSS supports the existence of three clusters, as it dramatically decreases from 4800.00 for one cluster to 1718.18 for three clusters, indicating better-defined groupings. This is in agreement with the Elbow method, which demonstrates that three clusters strike a balance between simplicity and model fit—that is, the clustering approach effectively isolates distinct student profiles. The findings align with those of scholars who have also applied clustering techniques to group different learner types, thereby demonstrating the effectiveness of clustering in discovering hidden patterns in educational data [16]. Further increases in clustering beyond three show increasingly marginal decreases in the within-cluster sum of squares, which vindicates the decision to limit the number of clusters to three for optimal results. This indicates that, through the use of K-means clustering on academic engagement and performance data, an effective way of grouping students with similar learning behaviors can be achieved, providing insights that guide personalized teaching strategies and interventions. The results also support learning analytics theory, which advocates for using clustering to segment students into meaningful groups, allowing prevailing interventions to cater to the specific needs of each group.

4. CONCLUSION

This study combines decision tree methodologies with K-means clustering techniques to predict academic success and identify distinctive learning patterns within the context of higher education in Bangladesh. By integrating a wide range of educational and behavioral indicators, this study develops an inclusive framework for understanding student success. The findings will help enhance learning analytics theory by demonstrating the importance of data-informed decision-making within educational organizations and in shifting toward a nuanced understanding of patterns that impact academic outcomes, thereby supporting targeted strategies aimed at improving student achievement. The findings demonstrate that it would be prudent for the institutions of higher education in Bangladesh to follow data mining tools that can help extensively in analyzing the academic and behavioral data of students. Improvements in data collection techniques, particularly about class attendance, submitted assignments, and online engagement, will result in more accurate predictions and individualized academic support. Customized interventions, based on student profiles—complete with classifications such as “active learners” and “at-risk students”—are expected to increase student engagement further, decrease the dropout rate, and raise overall academic achievement.

The present study is limited to using data from only one institution, which may limit the generalizability of the findings to other educational settings. Furthermore, the study focuses almost exclusively on academic and behavioral dimensions, while ignoring other factors that could influence student performance, such as socioeconomic status or other exogenous variables. Consequently, the precision and appropriateness of the model in other educational settings will require revalidation.

This study demonstrates the effectiveness of using decision trees and K-means clustering methods for predicting student outcomes and identifying distinctive learning patterns in higher education in Bangladesh. The study synthesizes critical academic and behavioral factors into a comprehensive framework to explain student success and inform targeted interventions. This research also progresses the domain of learning analytics and highlights the importance of data-driven decision-making. While this study has its limitations, it provides educators and policymakers with valuable insights into developing more effective academic support strategies, thereby facilitating improvement in students’ performance.

ACKNOWLEDGMENTS

We acknowledge the support from the university for helping us with the data.

FUNDING INFORMATION

This study did not receive any external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Md. Mahadhi Hasan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Md Nakibul Islam		✓				✓		✓	✓	✓	✓	✓		
Md Ikramul Haque Nirjon		✓	✓	✓		✓	✓		✓	✓	✓		✓	✓
Md Sharif Uddin			✓	✓		✓	✓		✓	✓	✓		✓	✓
Md. Muntasir Mamun					✓		✓		✓	✓			✓	✓
Zaheed Alam Munna					✓		✓	✓	✓	✓	✓			✓
Al Mahmud Rumman				✓	✓		✓			✓	✓		✓	✓

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest in this study.

ETHICAL APPROVAL

The research was carried out strictly according to ethical guidelines. In order not to violate privacy and to protect confidentiality, according to the laws of data protection, all information concerning students was anonymized. Informed consent for use of anonymized data was acquired from the university, and the data set did not include any personally identifiable information.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [MMH], on request.

REFERENCES

[1] M. M. Hossain, "English Language Teaching through Virtual Classroom during COVID-19 Lockdown in Bangladesh: Challenges and Propositions," *Journal of English Education and Teaching*, vol. 5, no. 1, pp. 41–60, 2021, doi: 10.33369/jeet.5.1.41-60.

[2] F. Rabbi, S. Islam, and J. Hossain, "Enhancing English Language Learning in Tertiary Education Through Blended Approaches: A Bangladesh Perspective," *Indonesian Journal of Education Research (IJoER)*, vol. 5, no. 1, pp. 1–9, Feb. 2024, doi: 10.37251/ijoer.v5i1.745.

[3] T. Tahmina, "Students' Perception of the Use of YouTube in English Language Learning," *Journal of Languages and Language Teaching*, vol. 11, no. 1, pp. 151–159, Jan. 2023, doi: 10.33394/jollt.v11i1.6883.

[4] S. A. Shorna and I. J. Suchona, "Utilizing Online Resources for Speaking and Listening Skills: Teachers' Perspectives in Bangladeshi Context," *Jurnal Bahasa Inggris Terapan*, vol. 10, no. 1, pp. 1–8, May 2024, doi: 10.35313/jbit.v10i1.5851.

[5] N. Sghir, A. Adadi, and M. Lahmer, "Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022)," *Education and Information Technologies*, vol. 28, no. 7, pp. 8299–8333, Jul. 2023, doi: 10.1007/s10639-022-11536-0.

[6] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, p. 11, Dec. 2022, doi: 10.1186/s40561-022-00192-z.

[7] C. T. McKenzie and D. A. C. Walma, "A cross-sectional study of academic success measures, grit, and resilience among US dental students," *Journal of Dental Education*, vol. 88, no. 11, pp. 1511–1519, Nov. 2024, doi: 10.1002/jdd.13627.

[8] A. S. N. Kim, S. Shakory, A. Azad, C. Popovic, and L. Park, "Understanding the impact of attendance and participation on academic achievement," *Scholarship of Teaching and Learning in Psychology*, vol. 6, no. 4, pp. 272–284, Dec. 2020, doi: 10.1037/stl0000151.

[9] A. Gambini, M. Desimoni, and F. Ferretti, "Predictive tools for university performance: an explorative study," *International Journal of Mathematical Education in Science and Technology*, vol. 55, no. 3, pp. 691–717, Mar. 2024, doi: 10.1080/0020739X.2021.2022794.

[10] M. O. Hamid and S. Sultana, "English-medium instruction in higher education in Bangladesh," in *The Routledge Handbook of English-Medium Instruction in Higher Education*, K. Bolton, W. Botha, and B. Lin, Eds., London: Routledge, 2024, pp. 352–364, doi: 10.4324/9781003011644-30.

[11] A. Al Masud, M. A. Hossain, S. Biswas, A. P. Ruma, K. S. Rahman, and S. Tagore, "The emergence of digital learning in higher education: a lesson from the COVID-19 pandemic," *The International Journal of Information and Learning Technology*, vol. 40, no. 3, pp. 202–224, May 2023, doi: 10.1108/IJILT-08-2022-0176.

[12] K. S. Ali and I. Hindarsah, "The higher education teaching practices in the public Universities of Bangladesh," *Journal of Economics and Business (JECOMBI)*, vol. 4, no. 2, pp. 82–90, 2024, doi: 10.58471/jecombi.v4i02.78.

[13] F. Amin and A. Y. Wahyudin, "The Impact of Video Game: 'Age of Empires Ii' Toward Students' Reading Comprehension on Narrative Text," *Journal of English Language Teaching and Learning*, vol. 3, no. 1, pp. 74–80, Jul. 2022, doi: 10.33365/jeltl.v3i1.1818.

[14] P. Dao, S. T. T. Hoang, and M. X. N. C. Nguyen, "Young learners' synchronous online peer interaction: teachers' beliefs of its benefits and implementation," *Language Awareness*, vol. 33, no. 1, pp. 45–69, Jan. 2024, doi: 10.1080/09658416.2022.2152828.

[15] J. Guryan *et al.*, "The Effect of Mentoring on School Attendance and Academic Outcomes: A Randomized Evaluation of the Check and Connect Program," *Journal of Policy Analysis and Management*, vol. 40, no. 3, pp. 841–882, Jun. 2021, doi: 10.1002/pam.22264.

[16] W. Ha, L. Ma, Y. Cao, Q. Feng, and S. Bu, "The effects of class attendance on academic performance: Evidence from synchronous courses during COVID-19 at a Chinese research university," *International Journal of Educational Development*, vol. 104, p. 102952, Jan. 2024, doi: 10.1016/j.ijedudev.2023.102952.

[17] S. M. Daily, M. L. Smith, C. L. Lilly, D. M. Davidov, M. J. Mann, and A. L. Kristjansson, "Using School Climate to Improve Attendance and Grades: Understanding the Importance of School Satisfaction Among Middle and High School Students," *Journal of School Health*, vol. 90, no. 9, pp. 683–693, Sep. 2020, doi: 10.1111/josh.12929.

[18] H. Silva, J. Lopes, G. Cruz, C. Dominguez, and E. Morais, "Does university attendance affect students' critical and creative thinking skills? A longitudinal research with pre-service teaching and psychology undergraduates," *Higher Education Research & Development*, vol. 42, no. 2, pp. 442–452, Feb. 2023, doi: 10.1080/07294360.2022.2057448.

[19] A. Akimov, M. Malin, Y. Sargsyan, G. Suyunov, and S. Turdaliev, "Student Success in a University First-Year Statistics Course: Do Students' Characteristics Affect Their Academic Performance?" *Journal of Statistics and Data Science Education*, vol. 32, no. 1, pp. 24–35, Jan. 2024, doi: 10.1080/26939169.2023.2184435.

[20] D. Hong, "Using in-class group assignments to improve low-achieving students' performance in an introductory accounting course," *Accounting Education*, vol. 33, no. 1, pp. 66–83, Jan. 2024, doi: 10.1080/09639284.2022.2119090.

[21] S. Conner, "How students leverage assignment submission flexibility — A case study," *Heliyon*, vol. 10, no. 21, p. e39937, Nov. 2024, doi: 10.1016/j.heliyon.2024.e39937.

[22] Z. K. Abdullaev, "Second language learning: Issues and implications," *Asian Journal of Multidimensional Research*, vol. 10, no. 4, pp. 824–827, 2021, doi: 10.5958/2278-4853.2021.00312.8.

[23] M. K. bin Chowdhury and F. P. Behak, "The blended learning as an alternative to the online education for the Bangladeshi universities," *International Journal of Language and Education*, vol. 3, no. 2, pp. 220–242, 2024.

[24] G.-J. Hwang, S.-Y. Wang, and C.-L. Lai, "Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics," *Computers & Education*, vol. 160, p. 104031, Jan. 2021, doi: 10.1016/j.compedu.2020.104031.

[25] M. K. Hossain and M. A. Al Younus, "Kahoot in the Classroom: A Game-Changer for Motivating English Grammar Learning in Tertiary Education in Bangladesh," *Journal of Critical Studies in Language and Literature*, vol. 5, no. 1, pp. 24–31, Jan. 2024, doi: 10.46809/jcsll.v5i1.251.

[26] M. H. Rahman, M. bin Amin, M. F. Yusof, M. A. Islam, and S. Afrin, "Influence of teachers' emotional intelligence on students' motivation for academic learning: an empirical study on university students of Bangladesh," *Cogent Education*, vol. 11, no. 1, p. 2327752, Dec. 2024, doi: 10.1080/2331186X.2024.2327752.

[27] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.

[28] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. Hoboken, NJ: Wiley, 2014.

[29] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and techniques*. Waltham, MA: Morgan Kaufmann, 2012.

[30] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks, 1984.

[31] R. L. Thorndike, "Who Belongs in the Family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, Dec. 1953, doi: 10.1007/BF02289263.

BIOGRAPHIES OF AUTHORS

Md. Mahadhi Hasan is an assistant professor, Department of English and Humanities, University of Liberal Arts Bangladesh. He obtained his Ph.D. in Communication (new media and language) from Infrastructure University Kuala Lumpur; BA from Comilla University; and an MA in TESOL from BRAC University. His research interests include new media and language, error analysis, AI in education, EdTech, indigenous literature, and South Asian literature. He is also interested in multidisciplinary research works. He has taught at BRAC University, Southeast University, Infrastructure University Kuala Lumpur, and East West University. He can be contacted at email: mahadhi4bd@gmail.com.

Md Nakibul Islam is a certified data analyst with extensive experience in marketing, finance, and sales management. He is currently pursuing a master's in IT with a focus on data management and analytics, while also engaged in a data analytics bootcamp at the University of Maryland Global Campus. With over two years of experience as a data analyst and nine years in sales and marketing, he has successfully applied data-driven strategies to exceed targets as a Territory Manager at Novartis Bangladesh Limited. His research focuses on utilizing data to generate valuable finance insights and enhance business decision-making processes. He can be contacted at email: nih644@gmail.com.

Md Ikramul Haque Nirjon is a dedicated and passionate individual with a strong background in technology and innovation. His journey from a young tech enthusiast to an aspiring data scientist is marked by academic excellence, professional growth, and a commitment to community service. He can be contacted at email: ikramnirjon@gmail.com.

Md Sharif Uddin holds a bachelor's in software engineering and a master's in IT with a focus on software design and management. He is passionate about the IT field, particularly data analytics, and aims to become a data expert. He can be contacted at email: mdsharif.iukl@gmail.com.

Md. Muntasir Mamun has 21 years of teaching experience at university level in English language and literature. He began his career as a lecturer at Eastern University, Dhanmondi, where he taught courses in both the honors and masters programs from 2004 to 2011. In addition to teaching there, he worked as the coordinator for the cultural club for more than four years. He has also taught at ASA University from 2011 to 2013. Muntasir joined ULAB's Department of English and Humanities as an assistant professor in 2013 and has been teaching here since. At ULAB, he serves as the advisor for the club theatre ULAB, and encourages and motivates students to become involved in cultural activities. He can be contacted at email: muntasir.mamun@ulab.edu.bd.

Zaheed Alam Munna joined Southeast University as a lecturer on 2006. Apart from teaching literature and language courses, he has been performing administrative responsibilities as the coordinator of the department. Recently, he has been appointed as the chairman of the department. Dr. Zaheed completed his BA (honors) in English Language and Literature and MA in English Literature from Jahangirnagar University. He was awarded Ph.D. degree in English Literature at University of Dhaka in November, 2024. Most of his research articles are on English romantic literature which is also his area of interest. So far, he has published 10 articles in peer-reviewed journals. He can be contacted at email: zaheed.munna@seu.edu.bd.

Al Mahmud Rumman has around eight years of experience in teaching at different levels of education. He is currently pursuing his MPhil degree on Discourse Analysis and Pragmatics from the Institute of Modern Languages, University of Dhaka. He completed his BA (honors) and MA in English from University of Chittagong and another MA in ELT from University of Dhaka. His teaching and learning interests include psycholinguistics, pragmatics, discourse analysis, sociolinguistics, and SLA. He can be contacted at email: mahmud.rumman@ulab.edu.bd.