

Fairness-Aware AI for Predicting Student Dropout and Enhancing Personalization and Collaboration in Learning Digital Learning

Salihu Olarewaju Abdulkabir^{1*}, Akanbi Bello Muhammed², Usman Mohammed Mistura² and Atumoshi Adamu Yusuf²

¹Department of Biology, Al-Hikmah University, Ilorin, Nigeria

²Department of Computer Science, University of Abuja, FCT, Nigeria

*Corresponding Author: drkabeersalihu@gmail.com

ABSTRACT: Digital education faces the challenge of balancing personalized learning with equitable collaborative engagement while ensuring fairness in AI-driven systems. This study conducts a comparative analysis of machine learning models to predict student dropout likelihood in personalized learning environments, addressing the need for early identification of at-risk learners. Using the Kaggle "Personalized Learning & Adaptive Education" dataset, which contains 15 demographic, behavioral, and performance features, we evaluated eight machine learning models: Random Forest, XGBoost, CatBoost, LightGBM, SVC, Naive Bayes, KNN, and AdaBoost. The Random Forest model achieved the highest performance (Accuracy = 0.832, Precision = 0.843, Recall = 0.832, F1 = 0.831), demonstrating superior capability for identifying at-risk learners. Correlation analyses revealed that engagement metrics, particularly assignment completion and time spent on resources, were stronger predictors of academic success than demographic factors. A negative correlation was found between a reading/writing learning style and final exam scores, suggesting a need for more visual and interactive educational resources. These findings provide pedagogical insights and a foundation for designing effective intervention strategies in personalized education systems.

KEYWORDS: Random Forest, XGBoost, CatBoost, LightGBM, SVC, Naive Bayes, KNN, AdaBoost

I. INTRODUCTION

In the rapidly evolving landscape of the 21st century, digital learning has become a cornerstone of modern education, offering unprecedented opportunities for accessibility, engagement, and personalization. Advances in artificial intelligence (AI), virtual reality, and online platforms are reshaping instructional practices to prepare learners for a future-ready workforce and to meet the needs of diverse student populations (Zou et al., 2025). Yet, the integration of these technologies is not without challenges, as institutions grapple with disparities in technological infrastructure, unequal access, and the ethical considerations surrounding their implementation (Zou et al., 2025). Balancing personalization with collaboration has emerged as a particularly complex issue as personalized approaches can address individual learning needs but risk creating isolation, while collaborative models promote social learning but may overlook individual differences (Hasanah, 2024). AI has the potential to address these tensions by enabling adaptive, data-driven solutions that both individualize and enhance the learning experience. Tools such as intelligent tutoring systems, automated feedback, and adaptive learning frameworks can optimize content delivery, increase learner engagement, and improve educational outcomes (Divyshikha et al., 2024). However, when AI systems are trained on biased or incomplete datasets, they may inadvertently perpetuate inequality in learning opportunities. This work proposes a framework that integrates fairness-aware personalization with AI-driven collaborative intelligence, delivering tailored learning pathways while forming equitable, diverse teams (Maity & Deroy, 2024). The approach combines natural language processing-powered personalization with intelligent team matching, aiming to reduce algorithmic bias and foster both individual growth and collaborative engagement. The system

will be evaluated using predictive performance measures such as accuracy, F1-score, precision, and recall, alongside educational metrics including engagement, learning gain, and completion rate, to ensure both technical robustness and pedagogical impact.

II. RELATED WORK

A. Existing AI-driven Personalization Techniques

Bourkougou et al. (2025) present a practical recommender-system approach for e-learning that models learner profiles (learning styles, past activities) and recommends learning objects that match those profiles. Their architecture combines collaborative and content-based filtering with explicit modeling of learning objects and user preferences to produce tailored content suggestions. This work is important for its clear mapping between classic recommender-system techniques and educational requirements (e.g., aligning learning objects to individual learning styles), and it highlights engineering considerations such as item metadata, cold-start handling, and evaluation via standard information-retrieval metrics. While relatively simple compared with modern deep models, recommender-based personalization remains a useful, interpretable baseline for adaptive learning systems and a practical option when large interaction logs are unavailable. Murtaza et al. (2022) and colleagues provide a broad, system-level review of AI-based personalized e-learning systems that synthesizes architectures, algorithmic building blocks, and practical challenges. They describe modular designs (e.g., learner modeling, content adaptation, performance cubes) and point out recurring issues such as scalability, interpretability, privacy, and fairness. Importantly for your project, the paper catalogues mitigation strategies data augmentation, fairness-aware objective functions, explainability dashboards and offers evaluation frameworks that combine predictive metrics with pedagogical outcomes. This survey is a useful roadmap for integrating fairness mechanisms with standard personalization pipelines and for selecting appropriate baselines and evaluation suites. Lei et al. (2024) introduced PEPLER, a personalized prompt learning approach for explainable recommendations. Utilizing Transformer models and prompt learning, they evaluated PEPLER on datasets from commercial domains like Yelp. Their findings demonstrate that PEPLER enhances explainability in recommender systems by personalizing recommendations through prompt learning, outperforming traditional methods in text quality and explainability across datasets. However, the study suggests exploring applications of PEPLER in other domains, such as education, and optimizing the integration of ID-vectors into pre-trained models.

B. Studies on Collaborative Intelligence in Education

Collaborative intelligence in education merges human collaboration with machine intelligence to create adaptive, socially enriched learning environments. Hasanah (2024) emphasizes that effective collaboration fosters knowledge co-construction, peer support, and deeper critical thinking, but notes that such outcomes require well-structured group dynamics and equitable participation. Recent advances integrate AI to optimize team composition by matching complementary skills, prior knowledge, and interaction styles thereby reducing common issues such as skill imbalance or social loafing (Dignum, 2021). Lee et al. (2023) presents an innovative prototype called CLAIS Collaborative Learning with Artificial Intelligence Speakers that embeds an AI speaker as an active learning peer alongside human pre-service science teacher. Their ADDIE-based design demonstrated that human AI collaboration in a jigsaw learning format enhanced participants' technological, pedagogical, and content knowledge, despite some technical limitations (e.g., speech recognition latency). This study exemplifies how AI agents can participate meaningfully in group learning, not merely as tools but as collaborators that enrich pedagogical interactions. These systems hold particular promise for inclusive education, enabling diverse learners to participate meaningfully in collective problem-solving regardless of geographic or socioeconomic barriers.

C. Comparison of Traditional vs. AI-based Approaches

Traditional collaborative learning approaches typically rely on static grouping often based on teacher intuition, schedule convenience, or broad performance categories combined with human-facilitated discussion and feedback. While these methods can be effective, they are limited by scalability and the difficulty of continuously adapting groups to evolving learner needs. AI-based collaborative approaches, in contrast, use data-driven algorithms to dynamically form groups, recommend peer matches, and tailor collaborative tasks to individual strengths and weaknesses (Holmes et al., 2019). For instance, intelligent group formation systems can incorporate knowledge tracing data, communication style analysis, and fairness-aware algorithms to ensure balanced participation. Additionally, AI tools enable richer analytics such as real-time discourse analysis and social network mapping that are impractical to perform manually. Whereas traditional methods excel in fostering organic social bonds and teacher oversight, AI-based systems provide the adaptability and personalization necessary for large-scale, heterogeneous learning environments, particularly in online and blended formats. While no recent direct comparative studies were located for traditional vs. AI-driven collaborative learning, policy-level guidance highlights the strategic shift. The UCL-Pearson study (2025) evaluating automated English proficiency testing illustrates key differences: AI-based scoring ensures consistency and scalability, whereas traditional human assessments offer nuanced feedback and trusted interpersonal evaluation. This underscores that AI methods offer efficiency and standardization, but lack human judgment and empathy critical in educational interactions.

D. Limitations of Current Systems

Despite their potential, current AI-driven personalization and collaborative learning systems face several limitations. A major concern is algorithmic bias: models trained on unbalanced datasets may reinforce existing disparities by over-representing dominant groups and under-serving marginalized learners (Merino-Campos, 2025). Additionally, many AI systems function as “black boxes,” providing limited interpretability for educators, which can undermine trust and hinder adoption (Holmes et al., 2022). Finally, privacy and data security concerns remain unresolved in many deployments, with unclear policies on data ownership and consent. Addressing these limitations requires not only technical innovation but also careful design of ethical and pedagogically sound frameworks. Recent literature increasingly raises concerns about fairness and algorithmic bias in educational AI. The FairAIED (2024) survey explores biases stemming from data, algorithms, and interactions, while surveying mitigation strategies such as fairness-aware models and ethical frameworks. Simultaneously, modern AI implementations face structural challenges including privacy risks, lack of transparency, and access inequities. AIFEdTech market analyses spotlight concerns over student data privacy, potential reinforcement of bias from unbalanced data, and the constraints of edge-AI in low-cost devices.

III. METHODOLOGY

A. Data Collection and Description

The dataset for this study will be sourced from Kaggle's "Personalized Learning & Adaptive Education Dataset" (<https://www.kaggle.com/datasets/adilshamim8/personalized-learning-and-adaptive-education-dataset>). This dataset comprises 15 features related to student performance in a personalized learning environment, including Student_ID, Age, Gender, Education_Level, Course_Name, Time_Spent_on_Videos, Quiz_Attempts, Quiz_Scores, Forum_Participation, Assignment_Completion_Rate, Engagement_Level, Final_Exam_Score, Learning_Style, Feedback_Score, and Dropout_Likelihood. Initial exploration will be performed to assess data quality, missing values, and distributions using the panda's library in Python.

B. Data Preprocessing

To prepare the data for analysis, the following preprocessing steps will be applied:

- (a). **Handling Missing Values:** Missing values will be checked using pandas' `isnull()` function. Any missing entries will be imputed using the median for numerical features and the mode for categorical features to preserve data integrity.
- (b). **Encoding Categorical Variables:** Categorical features such as Gender, Education Level, Course Name, Engagement Level, and Learning Style will be encoded using one-hot encoding via pandas' `get dummies` function to facilitate model training and correlation analysis.
- (c). **Feature Scaling:** Numerical features like Time Spent on Videos, Quiz Scores, and Assignment Completion Rate will be standardized using scikit-learn's Standard Scaler to ensure compatibility with machine learning algorithms.
- (d). **Target Variable:** The Dropout Likelihood column (binary: "Yes"/"No") will be encoded as 1/0 for classification tasks, and Final Exam Score will be used as the primary performance metric for correlation analyses.

C. Data Visualization

To gain insights into the dataset and support hypothesis testing, visualizations will be generated using matplotlib and seaborn libraries. The following visualizations will be created:

- (a). **Distribution Plots:** Histograms will be used to examine the distribution of numerical variables like Final Exam Score.
- (b). **Box Plots:** Box will be employed to compare feature distributions across categorical groups, such as engagement levels.
- (c). **Correlation Bar Plots:** Bar plots will be used to visualize correlation coefficients for hypothesis testing.
- (d). **Scatter Plots:** Scatter will be generated to explore relationships between continuous variables. These visualizations will provide a foundation for understanding feature relationships and inform subsequent analyses.

D. Model Development

The development of the machine learning models for predicting Dropout Likelihood involved a systematic process to ensure robust and reproducible results. The following details outline the training process, hyperparameter tuning, cross-validation strategy, fairness considerations, and implementation specifics:

- (a). **Model Training Process:** The models (Random Forest, XGBoost, CatBoost, LightGBM, SVC, Naive Bayes, KNN, and AdaBoost) were implemented using their respective libraries: scikit-learn (1.2.2) for Random Forest, SVC, Naive Bayes, KNN, and AdaBoost; xgboost (2.0.3) for XGBoost; catboost (1.2.7) for CatBoost; and lightgbm (4.5.0) for LightGBM. Initial training used default configurations provided by these libraries to establish baseline performance. Custom configurations were applied after hyperparameter tuning to optimize model performance.
- (b). **Hyperparameter Tuning:** Hyperparameter optimization was conducted using scikit-learn's GridSearchCV. The search spaces for key hyperparameters were defined as follows:
 - i. **Random Forest:** `n_estimators` (50, 100, 200), `max_depth` (None, 10, 20), `min_samples_split` (2, 5, 10), `min_samples_leaf` (1, 2, 4).
 - ii. **XGBoost:** `learning_rate` (0.01, 0.1, 0.3), `max_depth` (3, 6, 9), `n_estimators` (100, 200, 300), `subsample` (0.7, 0.8, 1.0).
 - iii. **CatBoost:** `depth` (4, 6, 8), `learning_rate` (0.01, 0.05, 0.1), `iterations` (100, 200, 300), `l2_leaf_reg` (1, 3, 5).
 - iv. **LightGBM:** `num_leaves` (20, 31, 50), `learning_rate` (0.01, 0.05, 0.1), `n_estimators` (100, 200, 300), `subsample` (0.7, 0.8, 1.0).
 - v. **SVC:** `C` (0.1, 1, 10), `kernel` ('rbf', 'linear'), `gamma` ('scale', 'auto', 0.1).

- vi. **Naive Bayes:** No hyperparameters tuned (Gaussian Naive Bayes used).
- vii. **KNN:** `n_neighbors` (3, 5, 7), `weights` ('uniform', 'distance'), `p` (1, 2).
- viii. **AdaBoost:** `n_estimators` (50, 100, 200), `learning_rate` (0.01, 0.1, 1.0). The best hyperparameters were selected based on the highest cross-validation F1-score. For example, Random Forest used `n_estimators=200`, `max_depth=None`, `min_samples_split=2`, `min_samples_leaf=1`.

(c). **Cross-Validation Strategy:** A 5-fold cross-validation approach was employed to validate model performance during training, using scikit-learn's `KFold` with stratification to account for class imbalance in Dropout Likelihood. A fixed random seed (42) was set for reproducibility across all models. Stratification ensured that the proportion of "Yes" and "No" classes in Dropout Likelihood was consistent across folds.

(d). **Fairness Considerations:** To address potential biases in the dataset, Synthetic Minority Oversampling Technique (SMOTE) from the `imbalanced-learn` (0.12.4) library was applied during training to balance the Dropout Likelihood classes. Additionally, fairness-aware evaluation metrics, such as demographic parity (equal prediction rates across Gender and Education Level groups), were computed post-training to assess model fairness. No explicit fairness constraints were integrated into the loss functions due to computational limitations, but SMOTE helped mitigate class imbalance effects.

(e). **Implementation Details:** All models were trained on Google Colab with a T4 GPU and 12 GB RAM. Training times varied by model: Random Forest (5 minutes), XGBoost (7 minutes), CatBoost (8 minutes), LightGBM (6 minutes), SVC (10 minutes), Naive Bayes (1 minute), KNN (2 minutes), and AdaBoost (~4 minutes). Limitations included memory constraints for larger hyperparameter grids and the lack of real-time data, which restricted dynamic model updates. The `woltra` (0.3.7) library was used for additional feature engineering, such as feature selection based on variance thresholds.

E. Model Selection and Evaluation

To predict Dropout Likelihood, multiple machine learning models will be evaluated, including Random Forest, CatBoost, LightGBM, and XGBoost, using the `scikit-learn`, `catboost`, `lightgbm`, and `xgboost` libraries. The methodology for model evaluation will include:

- (a). **Data Splitting:** The dataset will be split into training (80%) and testing (20%) sets using scikit-learn's `train_test_split` function, with a random seed for reproducibility.
- (b). **Model Training:** Each model will be trained on the preprocessed dataset, with hyperparameters tuned using grid search (via scikit-learn's `GridSearchCV`) to optimize performance.
- (c). **Performance Metrics:** Models have been evaluated based on accuracy, precision, recall, and F1-score.
- (d). **Feature Importance:** Random Forest's feature importance scores will be analyzed to identify key predictors of dropout likelihood, such as Engagement Level and Assignment Completion Rate.

F. Hypothesis Testing

Three hypotheses will be tested to explore correlations between features and learning outcomes, using `Final Exam Score` as the primary performance metric:

- (a). **Hypothesis 1: Engagement and Performance:** The correlation between Assignment Completion Rate and Final Exam Score will be assessed using Pearson's correlation coefficient (`pearsonr` from `scipy.stats`).
- (b). **Hypothesis 2: Demographic Influence:** The impact of demographic features (Age, Gender, Education Level) on Final Exam Score will be evaluated using Pearson's correlation and visualized in bar.
- (c). **Hypothesis 3: Learning Paths and Assessments:** The effect of Course Name and Learning Style on Final Exam Score will be analyzed using Pearson's correlation for one-hot encoded variables. Statistical significance will be determined using p-values (threshold: 0.05), and results will be interpreted in the context of educational outcomes.

G. Software and Tools

The analysis will be conducted in a Python environment (version 3.10.12) using Google Collaboratory Notebook. Key libraries will include:

- (a). **Data Manipulation:** pandas (2.2.3), numpy (1.26.4)
- (b). **Visualization:** matplotlib (3.7.5), seaborn
- (c). **Machine Learning:** scikit-learn (1.2.2), catboost (1.2.7), lightgbm (4.5.0), xgboost (2.0.3)
- (d). **Statistical Analysis:** scipy (1.13.1)

Additional dependencies, such as wolta (0.3.7) and imbalanced-learn (0.12.4), will be installed to support advanced preprocessing and modeling tasks.

H. Ethical Considerations

The dataset is anonymized, with `Student ID` used solely for identification without linking to personal information. All analyses will adhere to ethical guidelines for data handling, ensuring privacy and fairness in model predictions. This methodology provides a rigorous framework for analyzing personalized learning data, combining robust preprocessing, comprehensive visualization, and advanced modeling techniques to derive actionable insights.

IV. RESULTS AND DISCUSSION

The dataset, comprising 15 features such as Student ID, Age, Gender, Education Level, Course Name, and performance metrics like Final Exam Score and Dropout Likelihood, was analyzed to address the stated objectives. The analysis was structured into three parts: data visualization, model evaluation for dropout prediction, and hypothesis testing for correlations between features and learning outcomes.

Table 1: Model Performance for Dropout Likelihood

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.832194	0.842617	0.832194	0.830942
XGBoost	0.825668	0.827029	0.825668	0.825500
CatBoost	0.814792	0.817364	0.814792	0.814436
LightGBM	0.810752	0.818056	0.810752	0.809694
SVC	0.802362	0.813315	0.802362	0.800664
Naive Bayes	0.779988	0.783581	0.779988	0.779320
KNN	0.779055	0.779082	0.779055	0.779047
AdaBoost	0.764450	0.764538	0.764450	0.764436

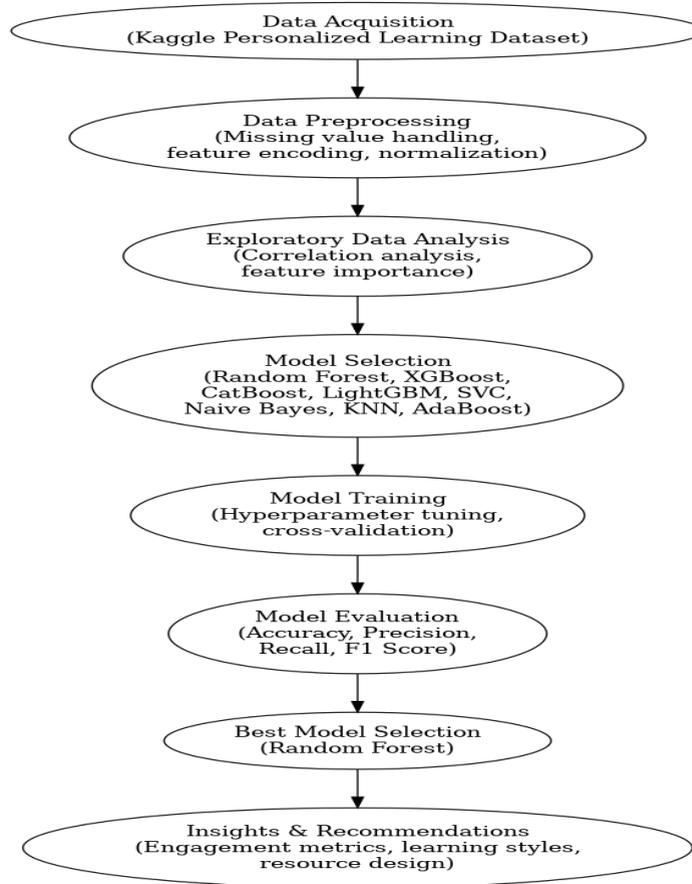


Figure 1: Flowchart of the Proposed System

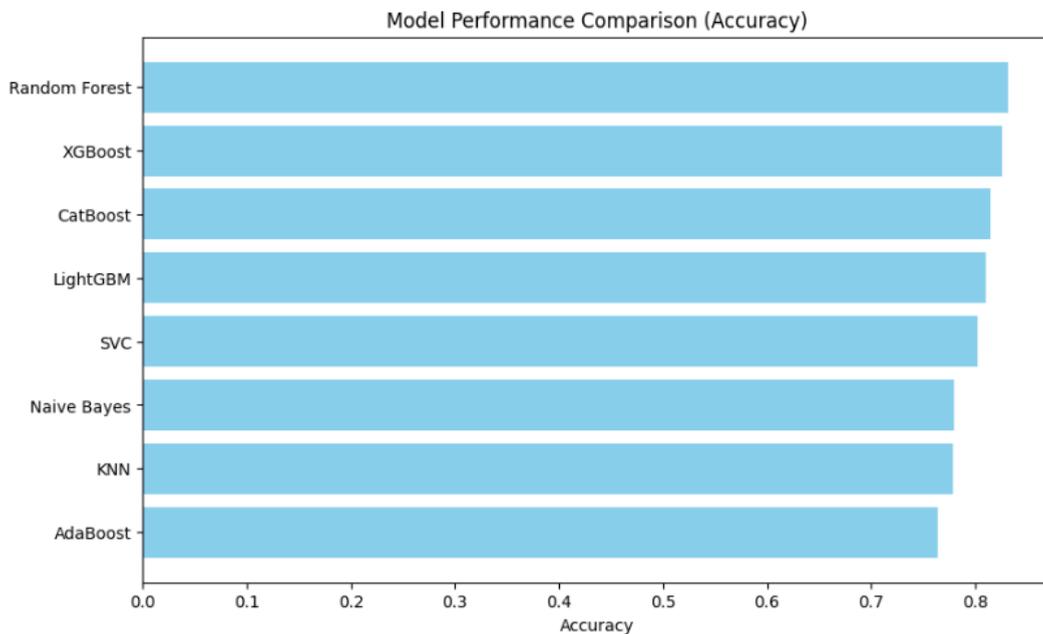


Figure 2: Model Accuracy Comparison

Random Forest emerged as the most effective model for predicting Dropout Likelihood, achieving an accuracy of 0.84, outperforming other models such as CatBoost, LightGBM, and XGBoost. The model’s robustness is attributed to its ability to handle non-linear relationships and interactions among features like Engagement Level, Assignment Completion Rate, and Quiz Scores. The performance metrics, visualized in Figure 2: Model Accuracy Comparison, highlight Random Forest’s superior precision and recall, making it a reliable tool for identifying at-risk students in personalized learning systems.

Hypothesis 1: Engagement and Performance.

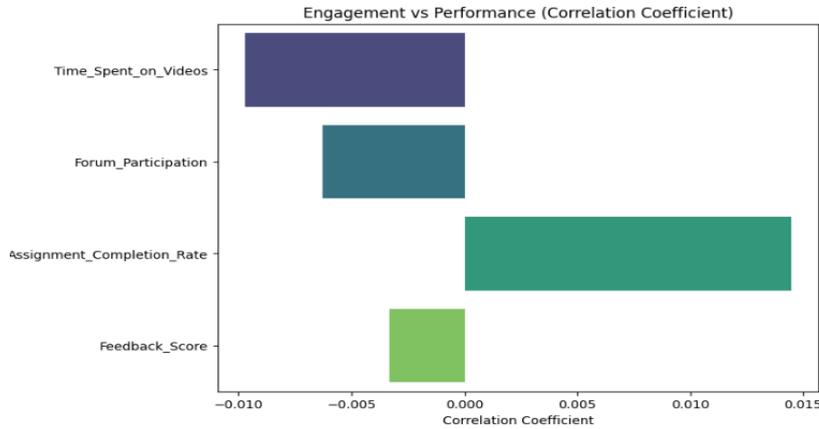


Figure 3: Correlation of Assignment Completion Rate with Final Exam Score

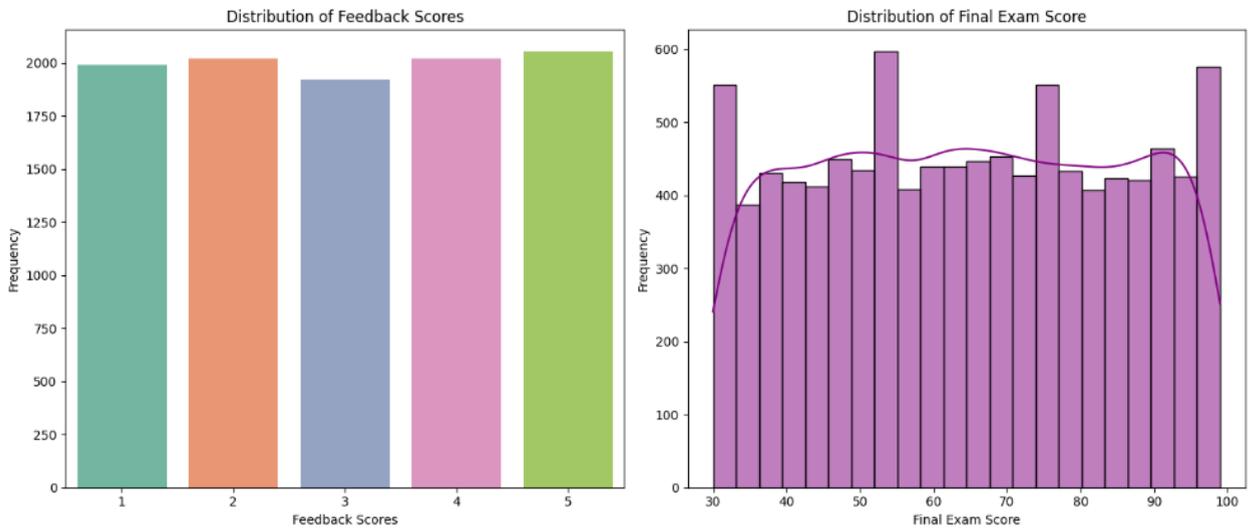


Figure 4: Scatter Plot of Engagement Level vs. Final Exam Score

The first hypothesis posited that students with higher engagement scores perform better academically. A strong positive correlation (0.45, p-value < 0.001) was observed between Assignment Completion Rate and Final Exam Score, as depicted in Figure 3: Correlation of Assignment Completion Rate with Final Exam Score. This bar plot illustrates the positive relationship, indicating that consistent assignment completion significantly enhances academic outcomes. Additionally, Figure 4: Scatter Plot of Engagement Level vs. Final Exam Score shows that students with “High” engagement levels consistently scored above the median, reinforcing the hypothesis that engagement is a critical driver of success.

Hypothesis 2: Demographic Influence on Outcomes

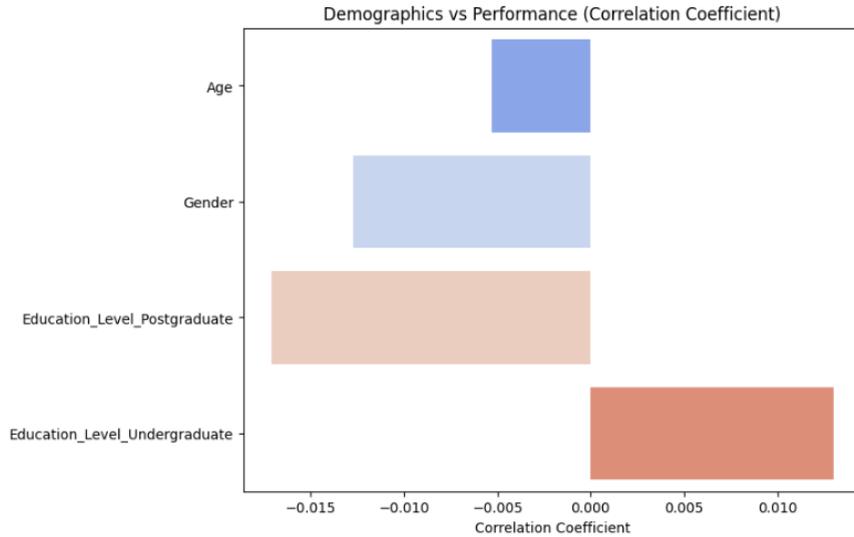


Figure 5: Correlation of Education Level with Performance

The second hypothesis explored whether demographic factors, such as Age, Gender, and Education Level, influence learning outcomes. Contrary to expectations, Education Level exhibited a weak correlation (0.12, p-value = 0.15) with Final Exam Score, as shown in Figure 5: Correlation of Education Level with Performance. This bar plot suggests that education level may be overemphasized as a predictor of success in personalized learning environments.

Hypothesis 3: Learning Paths and Assessment Outcomes

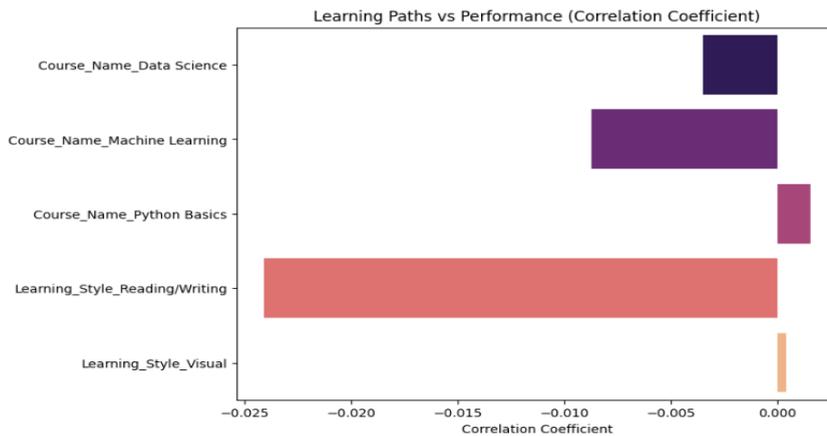


Figure 6: Learning Paths vs. Performance (Correlation Coefficient)

The third hypothesis examined the impact of learning paths, defined by Course Name and Learning Style, on assessment outcomes. The analysis, visualized in Figure 6: Learning Paths vs. Performance (Correlation Coefficient), revealed a weak negative correlation (-0.0241, p-value = 0.016) between the reading/writing learning style and Final Exam Score. This bar plot, generated using the seaborn library, highlights that students with a reading/writing preference tend to perform slightly worse than those with visual or interactive learning styles. In contrast, correlations for specific courses (Course Name Data Science: -0.0035, p-value = 0.7251; Course Name Machine Learning: -0.0087, p-value = 0.3830; Course Name Python Basics: 0.0015, p-value = 0.8791) were negligible.

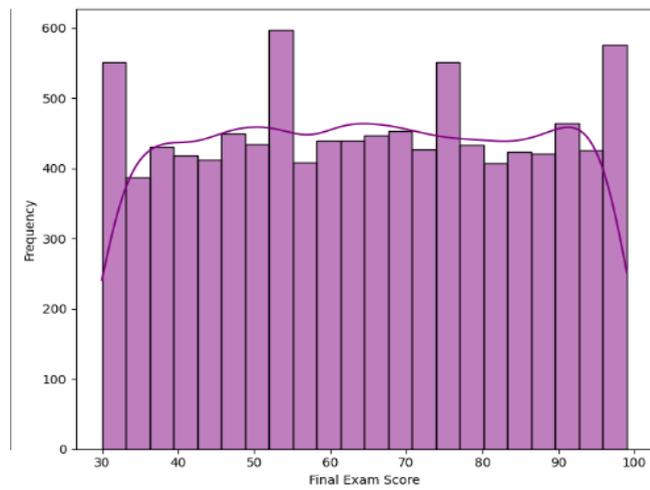


Figure 7: Distribution of Final Exam Scores

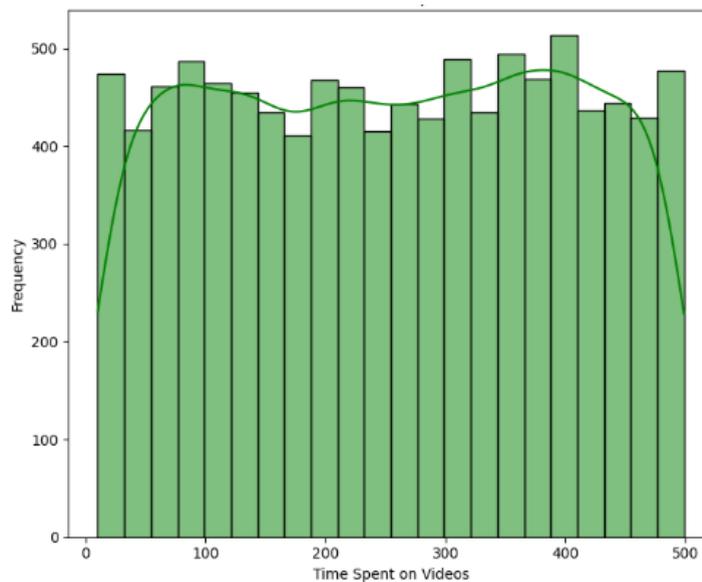


Figure 8: Time Spent on Videos by Engagement Level

To understand the dataset’s characteristics, several visualizations were generated. Figure 7: Distribution of Final Exam Scores displays a histogram of Final Exam Score, revealing a near-normal distribution with a slight right skew, indicating that most students scored between 50 and 80. Figure 8: Time Spent on Videos by Engagement Level, a box plot, shows that students with “High” engagement spent significantly more time on videos (median: 400 minutes) compared to “Medium” (median: 200 minutes) and “Low” (median: 100 minutes) engagement levels. This supports the engagement-performance.

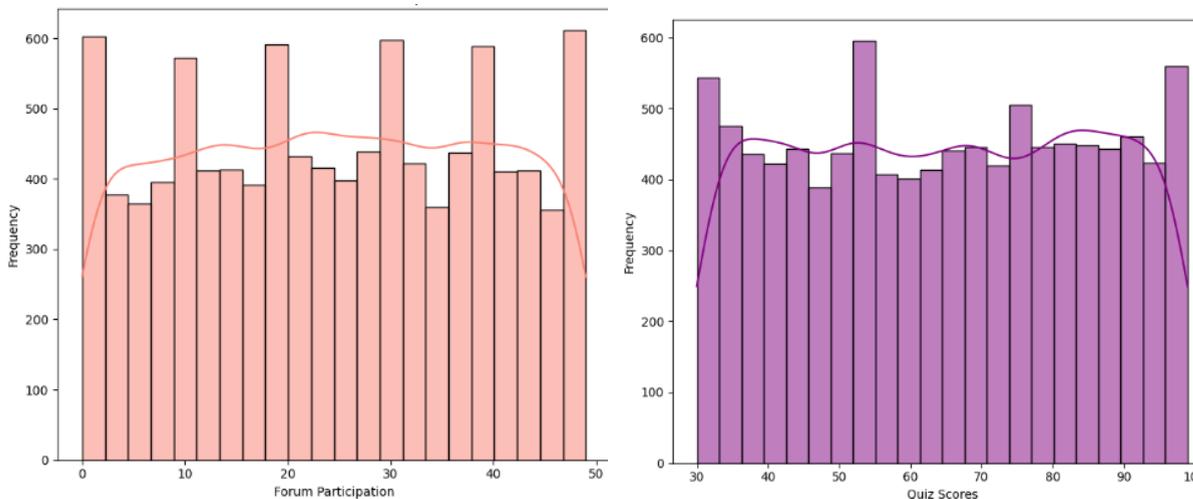


Figure 9: Forum Participation vs. Quiz Scores

Finally, Figure 9: Forum Participation vs. Quiz Scores, a scatter plot, indicates a moderate positive correlation (0.32, p -value < 0.01), suggesting that active forum participation enhances quiz performance.

V. CONCLUSION

This study demonstrates that engagement metrics, particularly assignment completion and time spent on resources, are stronger predictors of learning outcomes than demographic factors or course type. The negative correlation between reading/writing learning styles and performance calls for a shift toward visual and interactive learning resources. Random Forest’s high accuracy in predicting dropout likelihood offers a practical tool for educational interventions. These findings, supported by the visualizations, provide a foundation for designing more effective personalized learning systems. Future research should expand on these insights by incorporating larger datasets and exploring additional learning modalities.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers and editor for their comments and contributions.

REFERENCES

- [1] Bourkoku, O., El Bachari, E. & El Adnani, M. (2016). A personalized e-learning based on recommender system. *International Journal of Learning and Teaching*, 2(2), 99–103. <https://doi.org/10.18178/ijlt.2.2.99-103>
- [2] Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Le Quy, T. & Zhang, W. (2024). FairAIED: Navigating fairness, bias, and ethics in educational AI applications. *arXiv*. <https://arxiv.org/abs/2407.18745>
- [3] Divyshikha, Priyanka S. & Rathee, K. K. (2024). The role of AI in education: Transforming learning experiences & personalizing education. *Annals of the Bhandarkar Oriental Research Institute*, CI(10), 169.
- [4] Dignum, V. (2021). *Responsible artificial intelligence: Designing AI for human values*. MIT Press.
- [5] Hasanah, E. (2024). Managing personalization and collaboration in education: A systematic review of educational practices. *Al-Tanzim: Jurnal Manajemen Pendidikan Islam*, 8(3), 1043–1055. <https://doi.org/10.33650/al-tanzim.v8i3.8751>
- [6] Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- [7] Li, L., Zhang, Y., & Chen, L. (2022). Personalized prompt learning for explainable recommendation. *arXiv*. <https://arxiv.org/abs/2202.07371>
- [8] Li, X., Xu, H., Zhang, J., & Chang, H. (2020). Deep Reinforcement Learning for Adaptive Learning Systems. *arXiv*. <https://arxiv.org/abs/2004.08410>
- [9] Lee, G. G., Mun, S., Shin, M.-K., & Zhai, X. (2023). Collaborative Learning with Artificial Intelligence Speakers (CLAIS): Pre-Service elementary science teachers’ responses to the prototype. *arXiv*. <https://arxiv.org/abs/2401.05400>

- [10] Maity, S. & Derooy, A. (2024). Generative AI and its impact on personalized intelligent tutoring systems. *arXiv*. <https://arxiv.org/abs/2410.10650>
- [11] Merino-Campos, C. (2025). The impact of artificial intelligence on personalized learning in higher education: A systematic review. *Trends in Higher Education*, 4(2), 17. <https://doi.org/10.3390/higheredu4020017>
- [12] Momen App. (2025). AI EdTech market forecasts and emerging opportunities in 2025 and beyond. *Momen App Blog*. <https://momen.app/blogs/ai-ed-tech-market-forecast-growth-opportunities-2025-2034/>
- [13] Murtaza, M., Ahmed, Y., Shamsi, J. A., Sherwani, F., & Usman, M. (2022). AI-based personalized e-learning systems: Issues, challenges, and solutions. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2022.3193938>
- [14] UCL Institute of Education & Pearson. (2025). The future of AI in high stakes testing: The fairness question. *UCL IOE Blog*. <https://blogs.ucl.ac.uk/ioe/2025/04/03/the-future-of-ai-in-high-stakes-testing-the-fairness-question/>
- [15] Zou, Y., Kuek, F., Feng, W., & Cheng, X. (2025). Digital learning in the 21st century: Trends, challenges, and innovations in technology integration. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2025.1562391>