Students' Adaptability Level Prediction in Online Education using Interpretable Decision Tree based Model

Apeh Agene Gabriel
Department of computer Science
Federal University Lokoja, Kogi State, Nigeria

Taiwo Kolajo, PhD
Department of computer Science
Federal University Lokoja, Kogi State, Nigeria

Joshua Agbogun, PhD
Department of computer Science
Federal University Lokoja, Kogi State, Nigeria

ABSTRACT
This study investigates the prediction of students' adaptability levels in online education using machine learning, with a focus on decision tree–based algorithms. Guided by the objectives of dataset collection, framework design, model optimization, and evaluation, the study utilized secondary datasets sourced from existing research and an open university in the United States. Data preprocessing involved handling missing values, encoding categorical variables, scaling, and addressing class imbalance with SMOTE, followed by feature selection using RFECV. Four algorithms, namely C4.5-like Decision Tree, RepTree-like Decision Tree, RandomTree, and Random Forest, were optimized with Bayesian tuning and evaluated using accuracy, precision, recall, and F1-score. The results revealed RandomTree as the best-performing model (accuracy = 78.84%, F1 = 0.7892), a key finding that challenges the common superiority of Random Forest in related research. A generalizability assessment through nested cross-validation confirmed RandomTree's robustness (accuracy = 84.40% ± 0.0177 (mean/standard deviation)). Feature importance analysis highlighted age, financial condition, internet type, and location as critical determinants of adaptability. The findings emphasize the interplay of demographic, socio-economic, and technological factors in online learning adaptability, offering both theoretical contributions to educational data mining and practical tools for institutions to monitor and support students.

General Terms
Machine Learning, Educational Data Mining, Explainable Artificial Intelligence, Decision Tree Classification, Online Learning.

Keywords
Online Student adaptability, Students Adaptability Prediction Model, Interpretable Students' Adaptability Model, Decision Tree Based Students Adaptability Model.

1. INTRODUCTION
The COVID-19 pandemic abruptly shifted education from traditional classrooms to fully online environments, exposing significant differences in students' ability to adapt to remote learning. Understanding and predicting students' adaptability levels became essential to identify at-risk learners and to inform timely interventions that can improve engagement, learning outcomes, and educational resilience during and beyond pandemic-driven disruptions [1]

The ability of students to effectively adjust to digital learning environments which can be referred to as Student Adaptability in e-learning educational platform encompasses students' capacity to navigate technological platforms, manage self-directed learning, and cope with challenges such as reduced face-to-face interactions and increased autonomy. The adaptability of students is crucial in ensuring academic success, as students must develop strategies to stay motivated, regulate their time efficiently, and engage with course materials despite the absence of a structured physical classroom [2]

The shift to online education approach, specifically from the period of the COVID-19 pandemic to date, has highlighted the varying levels of adaptability among students. Factors such as prior exposure to digital tools, self-efficacy, internet accessibility, and socio-emotional resilience play significant roles in determining how well students adjust to virtual learning environments [1] Those with higher adaptability levels tend to exhibit stronger problem-solving skills, persistence, and proactive learning behaviors, while students who struggle with adaptation may face challenges such as procrastination, disengagement, and decreased academic performance.

Given the increasing reliance on online education, understanding and predicting students' adaptability has become essential for educators and policymakers. By identifying students at risk of poor adaptation, targeted interventions could be implemented to enhance learning experiences. The integration of machine learning models into this process allows for data-driven insights, making it possible to predict and address adaptability challenges in a timely and effective manner. Minh et al. [3]. Recent studies have explored the application of machine learning techniques to assess and predict student adaptability in online settings. For instance, Asencios-Trujillo et al. [4] utilized a Random Forest algorithm to classify students' adaptability levels based on demographic and contextual features. Similarly, Li [5] employed decision tree algorithms to analyze factors influencing students' adaptation to online education systems. These investigations highlight the efficacy of machine learning models in identifying key determinants of adaptability, such as age, educational background, and technological proficiency.

Moreover, the integration of Explainable Artificial Intelligence (XAI) techniques has been advocated to enhance the interpretability of these predictive models. Nnadi et al. [3] applied XAI methods to elucidate the contributions of various features affecting adaptability predictions, thereby providing actionable insights for personalized educational interventions. This approach facilitates a deeper understanding of the underlying factors influencing student adaptability, enabling educators to tailor support strategies effectively.

Motivated by previous studies this research aims to build upon these foundational studies by employing a comprehensive suite of machine learning approaches to predict students' adaptability levels in online education. By analyzing a diverse set of features this study seeks to develop robust predictive models that can inform the design of targeted interventions.

Existing model setbacks: Existing researches like Malik et al. [7] Suzan et al.[8] and Deepa and Mukesh [9] focused on predicting students' adaptability to online education using machine learning techniques, thus addressing the broader issue of assessing the effectiveness of digital learning environments, particularly in contexts where the transition to online education posed challenges. By analyzing students' socio-demographic factors and learning environments, the studies provided insights that can guide educational institutions in refining online education strategies. Various ML models were employed to classify students based on their adaptability levels. And ensemble models were consistently observed to achieve the highest accuracy, often outperforming simpler ML models. Some works also implement hyperparameter tuning to enhance model performance, showing that tuning can significantly improve predictive accuracy. Deepa and Mukesh, [9].

However, the common problems across these works are:

1. **Absence of optimizing the performance of all algorithms by tuning the hyperparameters:** While certain models such as Decision Tree in Deepa and Mukesh [9] work benefit from hyperparameter tuning, others (such as Naïve Bayes, KNN, SVM or even the ensemble models) are often applied without optimization. This results in potential performance gaps where tuning could have improved accuracy, precision, or recall for models.

2. **Limited generalizability due to dataset constraints** – Most studies rely on a single dataset, often sourced from a specific region or educational context. This may reduce the models' ability to generalize across diverse learning populations. Expanding data collection efforts and validating models across multiple datasets could enhance their real-world applicability.

3. **Low interpretability level of the existing models due to less integration of Explainable Artificial Intelligence (XAI) models into the problem domain:** Most studies rely on ML models that are less interpretable. This create a gap between the developed models with its applicability by the domain experts that are not familiar with ML domain.

To resolve these common problems, this study aim to develop a machine learning based model for predicting the adaptability level of students involved in online education. The focus during the model development will be to improve the robustness and generalizability of the machine learning. Hence, the contribution of this study to the domain of online student performance evaluation are:

1. Proposed an optimized Decision Tree based model for predicting students' adaptability level considering more than one online educational platform.
2. Develop an explainable/interpretable students' adaptability level prediction model. Thus, improving the understanding of education policy makers on machine learning solutions to online education problems relating to the student's adaptability.

2. RELATED WORKS

The integration of adaptability, online education, and machine learning represents a promising avenue for enhancing student outcomes in virtual learning environments. By leveraging ML algorithms, researchers can identify at-risk students and provide targeted interventions to improve their adaptability. For example, predictive models can be used to recommend personalized learning resources, adjust course designs, or provide timely feedback to students. Muthui et al, [10].

However, the successful implementation of ML in this context requires careful consideration of ethical and practical issues. Concerns about, algorithmic bias, and the interpretability of ML models must be addressed to ensure that these technologies are used responsibly and equitably. Additionally, the development of accurate and reliable predictive models depends on the availability of high-quality data and the selection of appropriate features (Baker, [11]). Based on these facts, this study focuses on reviewing existing researches related to this domain.

2.1 Decision Tree–Based Model Researches

Several studies have explored decision tree–based models due to their simplicity, transparency, and suitability for educational data analysis. Malik et al. [7] investigated students' adaptability in online entrepreneurship education by evaluating algorithms such as CART and C5.0 alongside other machine learning models using the Kaggle Educational dataset. Their findings demonstrated that decision tree–based models achieved competitive accuracy while maintaining interpretability, making them valuable for understanding how specific factors influence adaptability. Similarly, Suzan et al. [8] and Deepa and Mukesh [9] employed Decision Tree classifiers to categorize students according to adaptability levels based on socio-demographic and learning environment features. These studies showed that decision tree models effectively captured non-linear relationships in educational data, although their standalone performance was often surpassed by more complex models. Nonetheless, their transparent rule-based structure provided meaningful insights for educators seeking to identify key determinants of adaptability in online learning contexts.

2.2 Interpretable-Based Model Researches

Interpretable machine learning approaches have gained increasing attention in adaptability prediction due to the need for transparency and trust in educational decision-making. Nnadi et al. [6] presented a comprehensive study that integrated explainable AI (XAI) techniques with machine learning models to classify students' adaptability levels. Using Random Forest as the base classifier, the study applied SHAP, LIME, Anchors, ALE, and Counterfactual explanations to interpret model predictions. The results highlighted critical factors such as class duration and financial condition, while also demonstrating how small changes in feature values could significantly affect adaptability outcomes. In a related direction, Eid and Raju [12] emphasized interpretability by incorporating Copula-based dependency analysis into traditional machine learning models, including Logistic Regression and Decision Trees. Their approach enhanced the understanding of feature interactions and achieved high predictive accuracy, particularly with Logistic Regression, thereby reinforcing the value of interpretable models for actionable educational insights. Collectively, these studies underscore the importance of balancing predictive performance with model explainability in online education research.

2.3 Ensemble-Based Model Researches

Ensemble learning methods have consistently shown superior predictive performance in student adaptability studies by combining multiple learners to reduce variance and improve generalization. Malik et al. [7] demonstrated that ensemble-based models, particularly Random Forest, outperformed individual classifiers in predicting adaptability levels in online entrepreneurship education. Wang and Liu [13] further confirmed the effectiveness of ensemble techniques such as Random Forest, XGBoost, and CatBoost in modeling adolescent adaptability to online learning, identifying age, family financial status, and course duration as dominant predictors. Similarly, the studies by Suzan et al. [8] and Deepa and Mukesh [9] reported that ensemble models, including Random Forest and XGBoost, consistently achieved higher accuracy compared to simpler classifiers like Naïve Bayes and Logistic Regression. While these ensemble approaches delivered robust performance, they often introduced increased complexity, which limited interpretability despite the use of feature importance measures and tuning strategies to enhance results.

2.4 Summary and Research Gap

Overall, the reviewed literature demonstrates that machine learning techniques have been widely and successfully applied to predict students' adaptability in online education, especially following the COVID-19–driven shift to digital learning. Decision tree–based models offer transparency and ease of interpretation, interpretable machine learning approaches enhance trust and explanatory power, and ensemble models deliver high predictive accuracy. However, despite their strong performance, ensemble methods often trade model simplicity for accuracy, making them less suitable for stakeholders who require clear and actionable explanations. Additionally, issues such as limited model optimization and generalizability remain underexplored. These gaps highlight the need for research focused on less complex yet high-performing and interpretable machine learning models for predicting students' adaptability in online education.

3. METHODOLOGY

Based on the research framework designed in fig 1, the research framework for this study is structured as a systematic and iterative process designed to predict students' adaptability levels in online education using machine learning. It is guided by a conceptual framework that integrates data acquisition, preprocessing, model development, evaluation, and refinement. The framework ensures a structured pathway from raw data collection to the deployment of a reliable prediction model, with continuous validation and optimization embedded within.

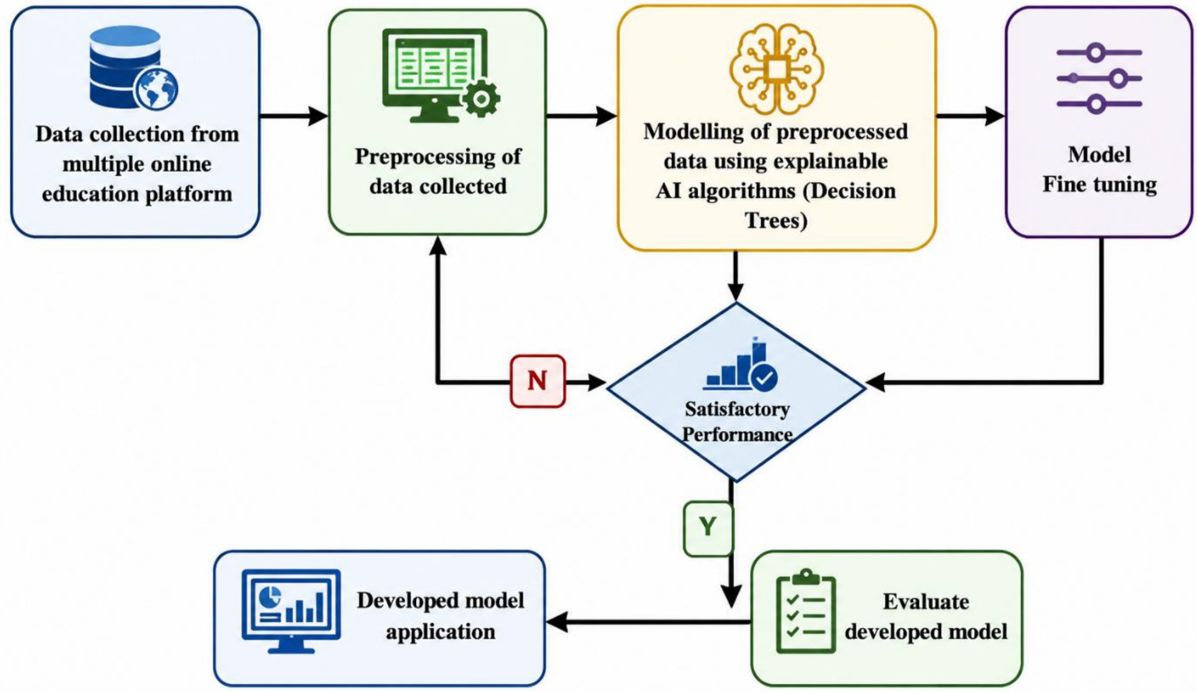

Fig 1 Study research framework

3.1 Data collection and preprocessing

The first phase of the framework involves data collection from multiple online educational platforms. These datasets encompass relevant features that reflect students' learning behaviors, engagement patterns, demographics, and academic records. This diversity in data sources helps to ensure that the model is generalizable and applicable across varied e-learning contexts. Following collection, the data undergoes a preprocessing stage, which includes handling missing values, normalization, encoding of categorical variables, and feature selection using . This step is crucial to prepare the data for accurate and efficient machine learning modeling.

3.1.1 Dataset

The data collection method for this study involves the use of secondary datasets, to be obtained from two main sources. The first dataset will be drawn from existing research conducted by Suzan et al.[8], while the second dataset is sourced from an open university in the United States. Both datasets represent student information from distinct online educational platforms and include variables relevant to students' engagement, demographic characteristics, academic performance, and learning behaviors. These datasets provide a diverse and rich foundation for training and evaluating the machine learning models used to predict students' adaptability levels in online learning environments.

The first dataset (as described in fig 2) was sourced from a public repository, comprising 1,205 entries with 14 categorical variables related to demographic, academic, and technological factors influencing dropout behavior. Key attributes included Gender, Age, Education Level, Institution Type, IT Student status, Location, Load-shedding impact, Financial Condition, Internet Type, Network Type, Class Duration, Self-LMS usage, Device type and the target variable Adaptability Level. All columns were complete, with no missing values, as indicated by the non-null counts matching the total entries.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1205 entries, 0 to 1204
Data columns (total 14 columns):
 #    Column                  Non-Null Count   Dtype
----  ------------------------ --------------  ------
 0    Gender                  1205 non-null    object
 1    Age                     1205 non-null    object
 2    Education Level         1205 non-null    object
 3    Institution Type        1205 non-null    object
 4    IT Student              1205 non-null    object
 5    Location                1205 non-null    object
 6    Load-Shedding           1205 non-null    object
 7    Financial Condition     1205 non-null    object
 8    Internet Type           1205 non-null    object
 9    Network Type            1205 non-null    object
 10   Class Duration          1205 non-null    object
 11   Self_Lms                1205 non-null    object
 12   Device                  1205 non-null    object
 13   Adaptivity Level        1205 non-null    object
dtypes: object(14)
memory usage: 131.9+ KB
```

Fig 2 Online repository dataset description (Suzan et al., [8])

The second dataset collected from Kaggle online repository was claimed to be obtained from a US Open University, containing 1,705 records with 8 features similar to the online dataset. Fig 3 depicts a description of the dataset. However, unlike the Suzan et al. [8] dataset, this dataset exhibited missing features (Load-shedding, Financial Condition and Self-lms). Both datasets were stored as Pandas DataFrames with most features encoded as categorical (dtype: object), necessitating preprocessing before modelling. For the purpose of generalizability both dataset were combined as depicted in fig 4, and by implication the missing features in the second dataset were removed from the first dataset. Without this, both dataset cannot be combined due to different number of features used in both dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1705 entries, 0 to 1704
Data columns (total 11 columns):
 #   Column                Non-Null Count   Dtype
---  --------------------  --------------   ------
 0   Education Level       1705 non-null    object
 1   Institution Type      1705 non-null    object
 2   Gender                1705 non-null    object
 3   Age                   1705 non-null    int64
 4   Device                1705 non-null    object
 5   IT Student            1705 non-null    object
 6   Location              1705 non-null    object
 7   Financial Condition   1705 non-null    object
 8   Internet Type         1705 non-null    object
 9   Network Type          1705 non-null    object
 10  Adaptability Level    1705 non-null    object
dtypes: int64(1), object(10)
memory usage: 146.7+ KB
```

Fig 3 A US Open University dataset description

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2910 entries, 0 to 2909
Data columns (total 11 columns):
 #   Column                Non-Null Count   Dtype
---  ----------------      --------------   -----
 0   Education Level       2910 non-null    object
 1   Institution Type      2910 non-null    object
 2   Gender                2910 non-null    object
 3   Age                   2910 non-null    object
 4   Device                2910 non-null    object
 5   IT Student            2910 non-null    object
 6   Location              2910 non-null    object
 7   Financial Condition   2910 non-null    object
 8   Internet Type         2910 non-null    object
 9   Network Type          2910 non-null    object
 10  Adaptability Level    2910 non-null    object
dtypes: object(11)
memory usage: 250.2+ KB
```

Fig 4 Combined dataset description

3.1.2 Data preprocessing

The dataset collected from online learning contexts underwent thorough preprocessing to ensure data quality and reliability. While missing values were not observed in the dataset, numerical features were standardized through scaling, and categorical variables were transformed via One-Hot Encoding (OHE) to make them machine-learning compatible. The dataset was then split into training (80%) and testing (20%) subsets of the combined dataset using a stratified strategy to preserve class distribution.

Class Imbalance Resolution: As observed in fig 5 the dataset class distribution was imbalance. To handle the issue of class imbalance in adaptability levels, SMOTE (Synthetic Minority Oversampling Technique) was applied to the training data. This ensured that minority classes were adequately represented, thereby improving the potential of the classifiers to recognize all adaptability categories fairly. The preprocessing stage established a robust foundation for subsequent modeling. After application the dataset class distribution became balance (as seen in fig 6) with equal number of class instances

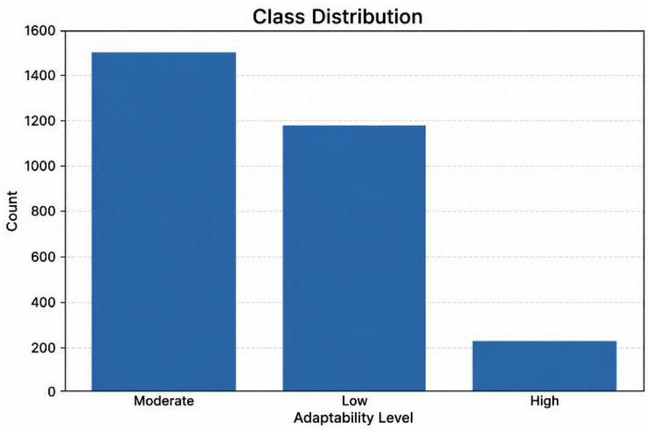

Fig 5 Dataset Class Distribution

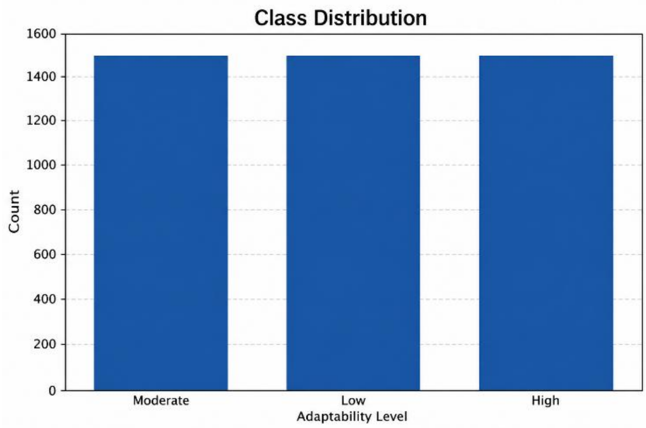

Fig 6 Dataset Class Distribution after SMOTE

Feature Selection: Feature selection was conducted using Recursive Feature Elimination with Cross-Validation (RFECV). This step was a preliminary step using a random algorithm (Random Forest) to reduce the dimensionality of the dataset by systematically eliminating irrelevant and redundant features while retaining those that provided the highest predictive value. However, all features were seen by the method to have significant values to predicting the adaptability of students in online learning platform.

However, to justify the RFECV results, a Forward Feature Selection method which focus on addition of important features recursively was adopted after the best model have been selected.

This method was also used to further enhance model interpretability for the best model by numerically explaining the significance of features to predicting students' adaptability

3.2 Modelling and Evaluation

The modeling phase leverages explainable artificial intelligence (XAI) techniques, particularly Decision Trees, to develop a predictive model. Decision Trees are chosen for their interpretability, allowing stakeholders such as educators and administrators to understand the basis of the model's predictions. Once the model is developed, it undergoes evaluation to determine whether it meets predefined performance criteria using standard metrics such as accuracy, precision, recall, and F1-score. If the model's performance is deemed unsatisfactory, it proceeds to the fine-tuning phase, where hyperparameters are adjusted and additional optimization strategies may be applied to improve results.

If the model achieves satisfactory performance, it is moved to the application phase, where it can be implemented for real-world use, such as identifying students at risk of poor adaptation to online learning environments and enabling timely interventions. The iterative nature of the framework where evaluation informs model refinement ensures that the final deployed model is both robust and effective in practical settings. This research framework not only supports the predictive goals of the study but also aligns with best practices in machine learning model development and evaluation.

3.2.1 Data analysis techniques

The data analysis in this study is centered on the application of multiple Decision Tree algorithms to develop an interpretable model for predicting students' adaptability levels in online education. The specific algorithms employed include C4.5, RepTree, RandomTree, and Random Forest. These algorithms are chosen for their ability to handle classification tasks effectively and provide varying levels of model interpretability and performance. The C4.5, RepTree, and RandomTree algorithms produce single-tree structures that allow for direct interpretation of decision paths, making it possible to explain how specific features contribute to the prediction.

However, due to the ensemble nature of the Random Forest algorithm, which builds multiple decision trees and aggregates their predictions, it does not produce a single interpretable model. To address this limitation and still derive meaningful insights, a wrapping technique using Forward Feature Selection is applied. This technique systematically evaluates the contribution of each feature to the model's performance by progressively adding features and observing the resulting changes in accuracy. This enables the identification and ranking of important predictive features, which is valuable for understanding the key factors influencing student adaptability in online learning environments.

Stepwise Algorithm Representation for Decision Tree Analysis

Step 1: Data Preparation

1. Load the preprocessed dataset.
2. Split the data into training and testing sets (e.g., 80% training, 20% testing).
3. Normalize or encode features as required by the algorithm.

Step 2: Apply Decision Tree Algorithm (Repeat for Each: C4.5, RepTree, RandomTree, Random Forest)

1. Initialize the model with default or optimized parameters.
2. Train the model using the training dataset.
3. Evaluate model performance using the testing dataset with metrics such as accuracy, precision, recall, and F1-score.
4. If applicable, visualize the decision tree (for C4.5 for the best performing model).
5. Record feature importance or structure for interpretability.

Step 3: Feature Importance Analysis (For the best performing model only)

1. Apply Forward Feature Selection by:
   a. Starting with an empty feature subset.
   b. Iteratively adding the feature that improves performance the most.
   c. Continuing until no significant improvement is observed.
2. Record the order and impact of selected features to determine key predictors.

Step 4: Comparison and Interpretation

1. Compare the performance across all algorithms.
2. Interpret the decision paths and key predictive features, especially from interpretable trees.
3. Use feature rankings from the best performing model with Forward Selection to support interpretability and model refinement.

Step 5: Model Generalizability assessment

1. Test the best performing model for its generalizability using the nested cross validation

4. RESULTS AND DISCUSSION

4.1 Results

The four-decision tree–based models were evaluated on the test dataset using accuracy, precision, recall, and F1-score as performance metrics. The results as observed in table 1 and plotted in fig 7, show slight variations among the algorithms, each reflecting its strengths in handling the classification task. The C4.5-like Decision Tree achieved an accuracy of 0.7759 with a corresponding F1-score of 0.7775, demonstrating a balanced performance though slightly limited by overfitting in deeper branches. The RepTree-like Decision Tree outperformed C4.5 with an accuracy of 0.7842 and F1-score of 0.7858, indicating that the pruning process improved its generalization ability. The RandomTree achieved the highest overall performance on the test set, with an accuracy of 0.7884 and F1-score of 0.7892, showing its ability to effectively capture non-linear patterns despite its randomized splitting process. By contrast, the Random Forest ensemble recorded a slightly lower accuracy of 0.7718 and F1-score of 0.7743, which, while robust, did not surpass the performance of the single-tree approaches in this dataset. Overall, the RandomTree emerged as the best-performing model in terms of test set predictive performance.

Table 1 Performance of the Decision Tree based models

Models	Accuracy	Precision	Recall	F1 Score
RandomTree	0.7884	0.8158	0.7884	0.7892
C4.5	0.7759	0.7921	0.7759	0.7775
RepTree	0.7842	0.8161	0.7842	0.7858
Random Forest	0.7718	0.8043	0.7718	0.7843

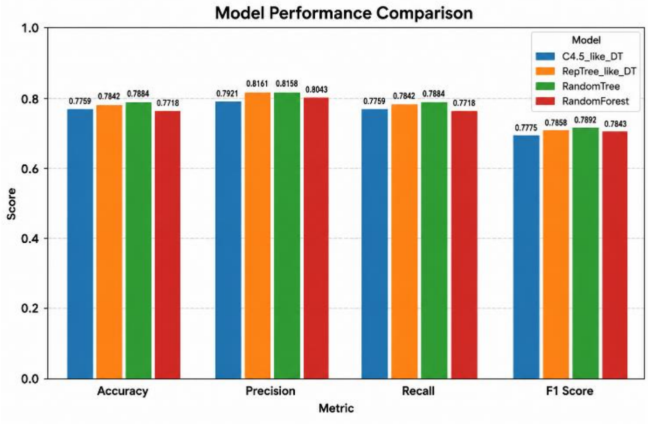

Fig 7 Comparison of Models Performances

To enhance interpretability, the RandomTree model, which achieved the highest predictive performance, was further analyzed to identify the most influential features in determining students' adaptability levels. The model's feature importance analysis as seen in fig 8 revealed the top ten splitting features ranked by Gini/Entropy measures. It should be noted that, every feature name used in the fig 8 is represented with the combination of the feature type and feature name, for example, feature Age is represented in fig 8 as num_Age (that is numeric Age). This helps distinguish what type of feature are ranked.

The most significant predictor was Age (0.1515), indicating that demographic factors played a strong role in adaptability prediction. Socio-economic indicators, particularly Financial Condition (Rich: 0.1343; Mid: 0.0486), were also highly influential, followed by Gender (Male: 0.0799; Female: 0.0406). Access and type of connectivity emerged as critical technical factors, with Internet Type (Mobile Data: 0.0720) and Network Type (4G: 0.0715) contributing substantially. Contextual factors such as Location (Rural: 0.0697) and Institution Type (Private: 0.0622) also played significant roles. Finally, the academic profile indicator IT Student (No: 0.0460) was among the top contributors.

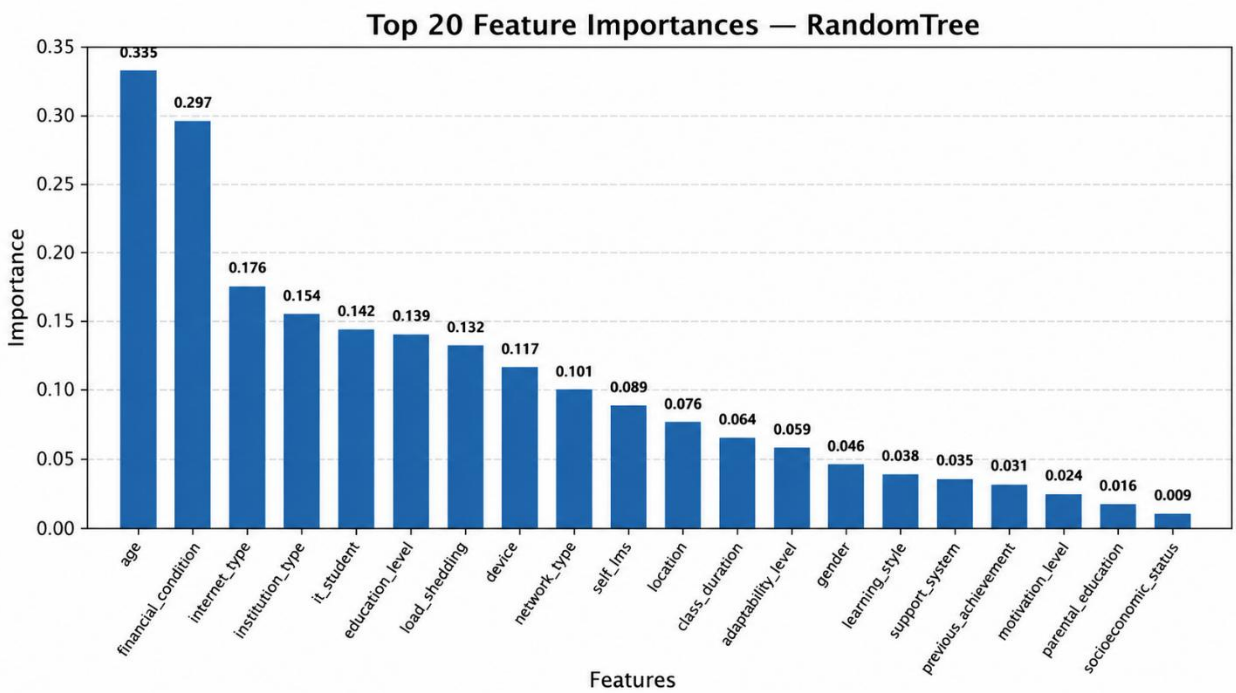

Fig 8 RandomTree Model Features Rankings

The generalizability of the best-performing RandomTree model was further assessed using nested cross-validation (CV) to ensure the stability and robustness of its performance across different data folds. The results demonstrated consistently strong outcomes as seen in table 2

Table 2 Generalizability of RandomTree (Nested CV)

Metrics	Generalizability values
Accuracy	0.8440 ± 0.0177
Precision	0.8440 ± 0.0188
Recall	0.8440 ± 0.0177
F1 Score	0.8418 ± 0.0184

4.2 Discussion

The results of this research provide valuable insights into the prediction of students' adaptability levels in online education using decision tree–based machine learning models. Aligned with the research objectives, the study successfully demonstrated the collection and preprocessing of domain-relevant data, the development of a predictive framework, the optimization of algorithms, and the evaluation of model performance.

Among the four decision tree–based algorithms evaluated—C4.5-like Decision Tree, RepTree-like Decision Tree, RandomTree, and Random Forest—the RandomTree model emerged as the best-performing algorithm on the test dataset, with an accuracy of

0.7884, precision of 0.8158, recall of 0.7884, and F1-score of 0.7892. Although Random Forest typically outperforms single-tree approaches, its test performance (accuracy = 0.7718, F1-score = 0.7743) was slightly lower in this study, emphasizing the importance of dataset characteristics in determining the best-suited algorithm.

Feature importance analysis further revealed that adaptability in online learning is influenced by a combination of demographic factors (e.g., age, gender), socio-economic conditions (e.g., financial status, institution type, location), and technological infrastructure (e.g., internet type, network connectivity). These findings provide a holistic view of the determinants of adaptability, offering critical insights for both researchers and educational practitioners.

The generalizability assessment of the RandomTree model through nested cross-validation confirmed its reliability, achieving an accuracy of 0.8440 ± 0.0177 and an F1-score of 0.8418 ± 0.0184. This result highlights the stability of the model across different subsets of the data, supporting its suitability for application in varied online education environments.

4.2.1 Comparative analysis with existing research

A comparative analysis was conducted between the results of this research and those reported by Suzan et al. (2021), who also employed machine learning approaches to predict students' adaptability in online education. In Suzan et al.'s study, the Random Forest algorithm emerged as the best-performing model, achieving an average accuracy of 90%, precision of 0.89, recall of 0.78, and F1-score of 0.83. These results set a strong benchmark for adaptability prediction in e-learning contexts.

In contrast, the findings of this research highlight RandomTree as the best-performing model, with an accuracy of 0.7884, precision of 0.8158, recall of 0.7884, and F1-score of 0.7892 on the test set. While the performance is robust and demonstrates the feasibility of decision tree–based approaches, the achieved scores are slightly lower than those reported by Suzan et al. (2021). Particularly, the accuracy in this study (78.84%) falls below the benchmark set by Suzan et al. (90%), while the F1-score (0.7892) is also marginally lower than the reported 0.83.

One key contribution by this study lies in the generalizability assessment of the best-performing model. Using nested cross-validation, the RandomTree model achieved an average accuracy of 0.8440 ± 0.0177 and F1-score of 0.8418 ± 0.0184, demonstrating stability across multiple folds. This suggests that, although the test set performance appears lower, the model exhibits consistent predictive ability and robustness, which is crucial for real-world deployment. By comparison, Suzan et al. (2021) reported only average performance scores without detailed cross-validation variance, leaving potential uncertainty about their model's stability across different subsets of data.

The deviation in results may be attributed to differences in datasets and feature composition. While Suzan et al. (2021) reported results on their proprietary dataset, this study combined secondary datasets from different online learning contexts, introducing heterogeneity in features such as demographic factors, financial conditions, and connectivity. This diversity, while improving the model's applicability to varied contexts, may have contributed to slightly reduced performance compared to a more uniform dataset. Additionally, the choice of RandomTree as the best-performing model, rather than Random Forest, deviates from Suzan et al.'s findings. This difference indicates that, in heterogeneous datasets, single-tree models with randomized splits can sometimes outperform ensemble approaches, especially when interpretability and feature-level insights are prioritized.

Nonetheless, this study's limitations include the lower accuracy relative to Suzan et al. (2021), highlighting the need for further optimization, possibly through hybrid or ensemble approaches that combine the strengths of multiple models. Future work could explore integrating advanced boosting techniques to bridge the performance gap.

Conclusively, while Suzan et al. (2021) achieved higher absolute performance scores with Random Forest, the present study contributes by demonstrating a robust and generalizable framework across diverse datasets and by identifying feature-level insights through RandomTree. This positions the study as complementary, extending the adaptability prediction literature by balancing predictive performance with interpretability and cross-context applicability.

5. CONCLUSION

This research has demonstrated the potential of decision tree–based machine learning algorithms in predicting students' adaptability levels in online education. By systematically collecting, preprocessing, and analyzing data, the study achieved its objectives and contributed both theoretical and practical insights. Theoretically, the research reinforces the importance of multi-dimensional predictors in adaptability modeling and challenges the assumption that ensemble models always outperform simpler approaches. The finding that RandomTree surpassed Random Forest in test performance suggests that dataset characteristics and the nature of feature distributions strongly influence model effectiveness. Furthermore, the generalizability demonstrated by nested cross-validation strengthens the confidence in deploying the model across different educational settings.

Practically, the research provides a predictive framework that institutions can adopt to monitor adaptability among students in real-time. The interpretability of decision tree models ensures that educators and administrators can trust and act upon predictions, enabling timely interventions such as additional training, financial support, or improved technological access for students likely to struggle. The identified features influencing adaptability such as financial condition, internet type, and location provide actionable insights for policymakers and educational planners to design targeted strategies that reduce barriers to online education. Although the study did not achieve the very high accuracy reported in related work by Suzan et al. (2021), its emphasis on interpretability and cross-context generalizability represents a significant contribution. In educational research, where transparency and reliability are critical, the balance achieved by this study between performance and interpretability makes the findings particularly valuable.

6. REFERENCES

[1] Bozkurt, A., and Sharma, R. C. 2023. Reimagining the Role of Generative AI in Distance Education and Online Learning. Asian Journal of Distance Education, 18(1), 1–7.

[2] Lovett, M. W., Frijters, J. C., Steinbach, K. A., Sevcik, R. A., and Morris, R. D. 2021. Effective Intervention for Adolescents with Reading Disabilities: Combining Reading and Motivational Remediation to Improve Outcomes. Journal of Educational Psychology, 113(4), 656.

[3] Minh, D., Wang, H. X., Li, Y. F., and Nguyen, T. N. 2022. Explainable Artificial Intelligence: A Comprehensive Review. Artificial Intelligence Review.

[4] Asencios-Trujillo, L. V., et al. 2025. Determining the Level of Adaptability of Students in Online Education Using Machine

Learning Algorithms. International Journal of Engineering Trends and Technology, 73(1), 305–312.

[5] Li, Y. 2024. Analysis and Prediction of Students' Adaptation to Online Education Systems Based on Data Analysis and Decision Tree Machine Learning Algorithms. Advances in Social Behavior Research, 7, 15–19.

[6] Nnadi, L. C., Watanobe, Y., Rahman, M. M., and John-Otumu, A. M. 2024. Prediction of Students' Adaptability Using Explainable AI in Educational Machine Learning Models. Journal of Artificial Intelligence in Engineering Practice, 1(2), 62–71.

[7] Malik, A., Onyema, E. M., Dalal, S., et al. 2023. Forecasting Students' Adaptability in Online Entrepreneurship Education Using Modified Ensemble Machine Learning Model. Array, 19, 100303.

[8] Suzan, M. H., Samrin, N. A., Biswas, A. A., and Pramanik, A. 2021. Students' Adaptability Level Prediction in Online Education Using Machine Learning Approaches. Proceedings of the 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1–7.

[9] Deepa, P. S., and Mukesh, K. R. 2023. Student Online Education Adaptability Prediction Using Machine Learning Algorithm. International Journal of Communication Networks and Information Security, 15(4).

[10] Muthui, P. M., Mwaura, P., and Mutua, J. N. 2023. Academic Adaptability as Predictor of Academic Achievement among Form Three Students in Kitui County, Kenya.

[11] Baker, R. S. 2024. Educational Data Mining and Learning Analytics. In Learning Analytics: From Research to Practice. Springer.

[12] Eid, M. M., and Raju, S. K. 2024. Predicting Student Adaptability in Online Education: A Comparative Study of Machine Learning Models and Copula-Based Analysis.

[13] Wang, M., and Liu, S. 2024. Machine Learning-Based Research on the Adaptability of Adolescents to Online Education.