Forum for Education Studies 2024, 2(3), 1415.
https://doi.org/10.59400/fes.v2i3.1415

Article

Enhancing adult learner success in higher education through decision tree models: A machine learning approach

Emily Barnes[1,*], James Hutson[2], Karriem Perry[3]

[1] Capitol Technology University, AI Center of Excellence (AICE), Laurel, MD 20708, USA
[2] Lindenwood University, Saint Charles, MO 63301, USA
[3] Capitol Technology University, Laurel, MD 20708, USA
*** Corresponding authors:** Emily Barnes, ejbarnes035@gmail.com

CITATION

Barnes E, Hutson J, Perry K. Enhancing adult learner success in higher education through decision tree models: A machine learning approach. Forum for Education Studies. 2024; 2(3): 1415. https://doi.org/10.59400/fes.v2i3.1415

ARTICLE INFO

Received: 3 June 2024
Accepted: 12 June 2024
Available online: 9 July 2024

COPYRIGHT

Copyright © 2024 author(s). *Forum for Education Studies* is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/by/4.0/

Abstract: This article explores the use of machine learning, specifically Classification and Regression Trees (CART), to address the unique challenges faced by adult learners in higher education. These learners confront socio-cultural, economic, and institutional hurdles, such as stereotypes, financial constraints, and systemic inefficiencies. The study utilizes decision tree models to evaluate their effectiveness in predicting graduation outcomes, which helps in formulating tailored educational strategies. The research analyzed a comprehensive dataset spanning the academic years 2013–2014 to 2021–2022, evaluating the predictive accuracy of CART models using precision, recall, and F1 score. Findings indicate that attendance, age, and Pell Grant eligibility are key predictors of academic success, demonstrating the strong capability of the model across various educational metrics. This highlights the potential of machine learning (ML) to improve data-driven decision-making in educational settings. The results affirm the effectiveness of Decision Tree (DT) models in meeting the educational needs of adult learners and underscore the need for institutions to adapt their strategies to provide more inclusive and supportive environments. This study advocates for a shift towards nuanced, data-driven approaches in higher education, emphasizing the development of strategies that address the distinct challenges of adult learners, aiming to enhance inclusivity and support within the sector.

Keywords: adult learners; decision tree models; machine learning; higher education; predictive accuracy

1. Introduction

In the landscape of higher education, the utility of machine learning (ML), especially Decision Tree (DT) models, has gained significant traction as a means to address the unique and multifaceted challenges encountered by adult learners [1,2]. Adult learners represent a substantial and growing segment of the student population, yet they face numerous socio-cultural, economic, and institutional obstacles that shape their educational experiences. These challenges include pervasive stereotypes that question their ability to learn, financial constraints that hinder their educational progress, and systemic inefficiencies within educational institutions that fail to meet their specific needs [3–6].

Despite the increasing enrollment of adult learners, there remains a significant gap in tailored educational strategies that address these unique challenges. The current study aims to fill this gap by exploring the application of ML, particularly DT models, to generate vital insights that can inform the development of policies, practices, interventions, and services designed specifically for adult learners. By leveraging the predictive capabilities of DT models, this research seeks to identify key factors that influence the success of adult learners, thereby providing educational institutions with data-driven tools to enhance support systems and improve educational outcomes for this demographic. The rationale behind this study is twofold. Firstly, it aims to address the underrepresentation of adult learners in educational research, particularly in the context of data-driven decision-making. Secondly, it seeks to demonstrate the efficacy of DT models in identifying and addressing the multifaceted challenges faced by adult learners. Through this approach, the study aspires to contribute to the development of more inclusive and effective educational strategies that can accommodate the diverse needs of adult learners, ultimately fostering a more equitable higher education landscape.

Adult learners represent a substantial and growing segment of the student population, yet they face numerous socio-cultural, economic, and institutional obstacles that shape their educational experiences. These challenges include pervasive stereotypes that question their ability to learn, financial constraints that hinder their educational progress, and systemic inefficiencies within educational institutions that fail to meet their specific needs [3–6]. The complexity of these challenges is further exacerbated by issues such as inter-role conflict, the need for education that adapts to changing job markets, and the necessity to balance academic pursuits with work and family responsibilities. In this evolving educational context, machine learning emerges not only as a tool to understand but also to effectively address the needs of adult learners by facilitating the creation of responsive and inclusive educational strategies. DT models, particularly Classification and Regression Trees (CART), can identify patterns and predictors of student success by analyzing vast amounts of educational data. This capability allows institutions to tailor interventions that directly address the specific barriers faced by adult learners.

For example, DT models can highlight how factors such as attendance, age, and Pell Grant eligibility influence graduation outcomes. By understanding these relationships, educational institutions can develop targeted support systems that enhance student engagement, provide financial aid counseling, and offer flexible scheduling options to accommodate work and family responsibilities. Moreover, these models can help predict potential dropouts, allowing for timely interventions that support at-risk students.

Furthermore, with adult learners now representing a substantial segment of the student population in the United States, there is a pressing need to adapt educational offerings, services, and practices to ensure equitable participation in contemporary educational environments [7–9]. This article delves into the application of machine learning within higher education frameworks, assessing the effectiveness of CART decision tree models in predicting graduation outcomes for adult learners. By leveraging these predictive models, this research aims to provide actionable insights that can inform policy and practice, ultimately contributing to more inclusive and effective educational strategies for adult learners.

The DT model, a staple in the realm of data mining, sufficiently handles classification and prediction tasks, making it a valuable tool for educational data analysis [10]. This model distinguishes itself through its recursive approach, which simplifies complex decision-making processes into understandable and easily interpretable models. Such clarity is particularly advantageous when evaluating graduation rates among adult learners in postsecondary education, where it is important to make informed predictions about student educational trajectories [10,11]. Its proficiency in leveraging historical academic performance to forecast future educational outcomes makes the DT model particularly suited for this task [12,13]. The model's effectiveness extends to classifying and predicting levels of student achievement with notable precision, utilizing algorithms such as the Iterative Dichotomiser 3 (ID3) to predict graduation outcomes from early academic indicators [14]. Further developments, such as improvements to the C4.5 DT algorithm, an effective technique for creating decision trees that can produce rules from the tree and handle both discrete and continuous attributes, have significantly enhanced its predictive accuracy, making it more reliable for classifying and forecasting student performance across various educational metrics [10,14].

2. Literature review

The utility of Classification and Regression Trees (CART) in higher education analytics is underscored by their efficiency in dissecting complex categorical data. Their straightforward and interpretable structure facilitates initial analysis and feature selection, making them ideally suited for examining the influence of various attributes, including categorical ones. The foundational work on Decision Trees (DT), based on Quinlan's [15] C4.5 algorithm, illustrates their application in predicting the success rates of adult learners, providing a robust framework for subsequent studies.

2.1. Graduation rates and academic performance

In the realm of academic outcomes, Gotardo [16] successfully utilized decision trees (DT) to predict graduation rates and licensure examination success with notable precision. Similarly, Alsariera [17] identified artificial neural networks (ANNs) as highly effective in forecasting student performance, a finding supported by Martins et al. [18] and Garg et al. [19], who noted the superior predictive accuracy of such networks. Tarmizi et al. [20] and Ivanov [21] further demonstrated the efficacy of decision trees in assessing mathematical competencies and predicting academic achievement, highlighting the strong predictive capacity of these models. Recent studies have continued to affirm the utility of ML models in educational settings. Crismayella et al. [22] employed the Adaboost algorithm alongside DT to classify graduation rates, achieving an accuracy increase from 70% to 82.14% through boosting. Nguyen et al. [23] demonstrated that RF models could effectively predict on-time graduation, identifying academic processing information and GPA as critical predictors. Additionally, the team found that multiple ML algorithms, including DT, could reliably forecast graduation outcomes based on various academic and demographic factors.

2.2. Analyzing academic and employment outcomes

Akmeshe et al. [24] and Zhou et al. [25] validated the efficiency of DT in predicting academic performance, particularly when analyzing e-learning behaviors and socio-demographic factors. Oreški et al. [26] applied the DT algorithm to a

Croatian dataset, achieving high accuracy in predicting student outcomes, while Khor [27] found decision trees superior in the early detection of low-performing students, emphasizing the importance of assessment scores and virtual learning interactions. Furthermore, Yan et al. [28] employed DT to predict student employment outcomes based on educational backgrounds and work experiences. Recent studies have further expanded on these findings. Meng [29] utilized DT to analyze and predict the employment status of college students, finding significant factors that affect student employment outcomes. Similarly, Pradana et al. [30] compared different tree-based algorithms and identified the Logistic Model Tree (LMT) as the most accurate for predicting employee attrition. Yin [31] proposed an improved DT method for optimizing employment guidance, demonstrating superior performance in improving employment rates and internship rates compared to traditional methods.

2.3. Mental health assessment and model versatility

Muzumdar et al. [32] demonstrated the eXtreme Gradient Boosting (XGBoost) model's superior efficacy in assessing student mental health, showcasing the potential of advanced ensemble techniques to identify patterns and predictors of mental health issues, thereby enabling timely interventions. Alsariera et al. [17] and Teng et al. [6] highlighted the effectiveness of support vector machines, DT, and Random Forests (RF) across various educational applications, from predicting graduation rates to facilitating data-driven decision-making. Recent studies have further underscored the versatility of DT in mental health assessment. Battista et al. [33] applied decision trees to youth mental health survey data, finding them particularly useful for identifying high-risk subgroups for targeted intervention. Li [34] improved the traditional ID3 algorithm for mental health assessment, enhancing classification accuracy and model fit. Additionally, Xiaocheng [35] used DT to evaluate college student mental health, specifically predicting depression, anxiety, and suicidal ideation. Furthermore, Krishnan et al. [36] applied DT alongside support vector machines to predict mental health issues in higher education students during the COVID-19 pandemic, achieving high classification accuracy.

2.4. The role of machine learning in higher education

The integration of machine learning in higher education transcends traditional frameworks, playing a crucial role in the analysis of student performance data and forecasting educational outcomes. Various ML algorithms, including RF, DT, and NN, have proven effective in forecasting student achievement, retention, and graduation probabilities. This body of research emphasizes ML's capacity to enhance data-driven decision-making and support the development of inclusive and flexible educational models that cater to adult learners [3,21,37,38]. Recent studies continue to highlight the transformative impact of ML in higher education. Tahiru et al. [39] conducted a bibliometric analysis illustrating the global research trends in ML-based predictive systems in higher education, emphasizing the growing contributions from countries like China, Belgium, and Spain. Mao, Fang, and Xia [40] developed a model using educational AI to predict the influence of student participation on final exam results, showcasing how ML can improve teaching methods in real-time. Pinto et al. [41]

conducted a systematic literature review identifying the prediction of academic performance and employability as the most researched applications of ML in higher education.

This literature review explores the role of ML in enhancing educational outcomes for adult learners. It highlights the diverse applications of ML, from predictive analytics to the creation of adaptive learning systems, discussing their potential to meet the challenges faced by adult learners in higher education. The review identifies key areas where machine learning has been applied, such as student performance prediction and dropout risk assessment, illustrating how these applications contribute to more engaging and effective learning environments [6,42]. This synthesis points to a continued need for research to refine these models, enhance their accuracy, and expand their applicability, allowing institutions to fully leverage the benefits offered by ML technologies.

3. Methods

In this study, a quantitative evaluation of CART DT models was conducted to assess their effectiveness in predicting degree completion rates among adult learners. Utilizing key statistical measures such as accuracy, precision, recall, and F1 score, the analysis aimed to determine which model provided the most reliable forecasts of academic success. The dataset encompassed variables including age, ethnicity, gender, Pell Grant eligibility, and academic performance metrics, spanning the academic period from 2013–2014 to 2021–2022. This comprehensive dataset allowed for a robust analysis of various factors impacting academic outcomes among adult learners.

3.1. Data collection and preprocessing

The initial stages of the study involved data collection and preprocessing, which are crucial for maintaining the integrity and confidentiality of the data. Data was collected securely from a student management system (SMS), anonymized, and stored in a cloud-based system with stringent security measures to prevent unauthorized access. The preprocessing phase involved rigorous data cleaning to correct inaccuracies, integration of data from various sources, and transformation techniques such as normalization and encoding. These steps were essential for preparing the data for machine learning algorithms, ensuring that the data was accurate, cohesive, and suitable for detailed analysis [43].

The preprocessing efforts also included the validation and transformation of data to ensure uniformity and reliability for analysis. Issues such as missing values were addressed by removing records, refining the dataset to 9999 records with enhanced reliability. The feature selection process was driven by a deep understanding of how various demographic, socioeconomic, and academic factors influence educational pathways. Features such as student type, generation, gender, age, ethnicity, and Pell Grant eligibility were considered for their potential impact on educational trajectories. This approach to feature selection and data preparation aimed to optimize the predictive accuracy of the models and provide meaningful insights that could inform educational strategies and support systems for adult learners [33,44].

3.2. Model and evaluation

The model building and evaluation procedures for this study involved a structured approach, beginning with the segmentation of the dataset into training and testing subsets. An 80/20 split ratio was employed, assigning 7999 entries to the training subset and 2000 entries to the testing subset. This division is a widely recognized strategy within ML and data science communities, as it ensures a balanced allocation for model training and validation [45]. The training subset is used to calibrate the model's parameters and allows the model to learn from a comprehensive array of data points and scenarios, which is critical for understanding the underlying patterns within the data. Conversely, the testing subset is used to evaluate the model's performance on new, unseen data, assessing its ability to generalize and maintain accuracy across different scenarios.

The development of the predictive models specifically focusing on CART followed a detailed and recursive methodology. The initial parameter settings for this were selected based on a combination of industry best practices and empirical research findings. The initial parameters were chosen to prevent overfitting, especially in highly flexible models such as GBM and DT. The initial parameter selection for each model was a deliberate process guided by established ML theories, empirical evidence from the literature, and insights gained from our dataset's preliminary analysis [43].

This approach ensured that the model (**Table 1**) was well-suited to uncover meaningful patterns and relationships within the educational data, setting a strong foundation for the subsequent iterative optimization process. Further, this method aimed to optimize the models' predictive accuracy and generalization capabilities, ensuring that they performed consistently well across various data samples. This iterative refinement and evaluation phase is essential for fine-tuning the model to achieve the highest level of accuracy and reliability in predicting outcomes [46].

Table 1. Decision tree model parameters.

Model	Max_Depth	Min_Samples_Split
DT1	5	2
DT2	3	3
DT3	2	3

Note. **Table 1** outlines the parameters and accuracy for three separate iterations of the decision tree model—DT1, DT2, and DT3.

3.3. Findings

The CART models, designated DT1, DT2, and DT3, have been rigorously evaluated for their capability to classify and predict educational outcomes, with the findings presented in **Tables 2** and **3**.

Table 2. Decision tree performance metrics.

Model	Accuracy	Mean CV Score	Standard Deviation
DT1	0.8315	0.60008	0.143494
DT2	0.8315	0.70128	0.145324
DT3	0.8085	0.65158	0.162175

Note. Number of records in training set = 7999, number of records in testing set = 2000.

Table 3. Decision tree model effectiveness and reliability scores.

Model	Accuracy	Precision	Recall	F1-Score
DT1	0.8315	0.815160	0.954086	0.879168
DT2	0.8315	0.795143	0.993774	0.883431
DT3	0.8085	0.897707	0.792218	0.841670

Note. Number of records in training set = 7999, number of records in testing set = 2000.

The performance metrics for the DT models (DT1, DT2, and DT3) indicate varying levels of accuracy and generalizability. Both DT1 and DT2 models achieved an accuracy of 83.15 percent, demonstrating strong predictive capabilities. However, DT2 stands out with the highest mean cross-validation (CV) score of 70.12 percent, indicating superior generalizability across different subsets of the dataset. Despite having a slightly higher standard deviation than DT1, DT2 shows the least variability in performance among the three models. On the other hand, DT3, although displaying a lower accuracy rate of 80.85 percent, maintains a competitive edge with a mean CV score higher than that of DT1. The relatively high standard deviation of 0.162175 for DT3 suggests more variability in its performance across various data samples, pointing to a potential need for further optimization to enhance its predictability and stability.

The results from **Tables 2** and **3** demonstrate the effectiveness of CART DT models in educational data analysis. DT2, in particular, stands out with its high accuracy and impressive recall score, highlighting its capability to accurately predict outcomes and reliably identify relevant cases. The overall performance of the DT models underscores their utility in classifying and predicting educational outcomes, with each model offering distinct advantages for different application needs. Additionally, **Figure 1** showcases a confusion matrix that visualizes the actual versus predicted values for DT1, DT2, and DT3, further illustrating the potential of these models in educational settings.

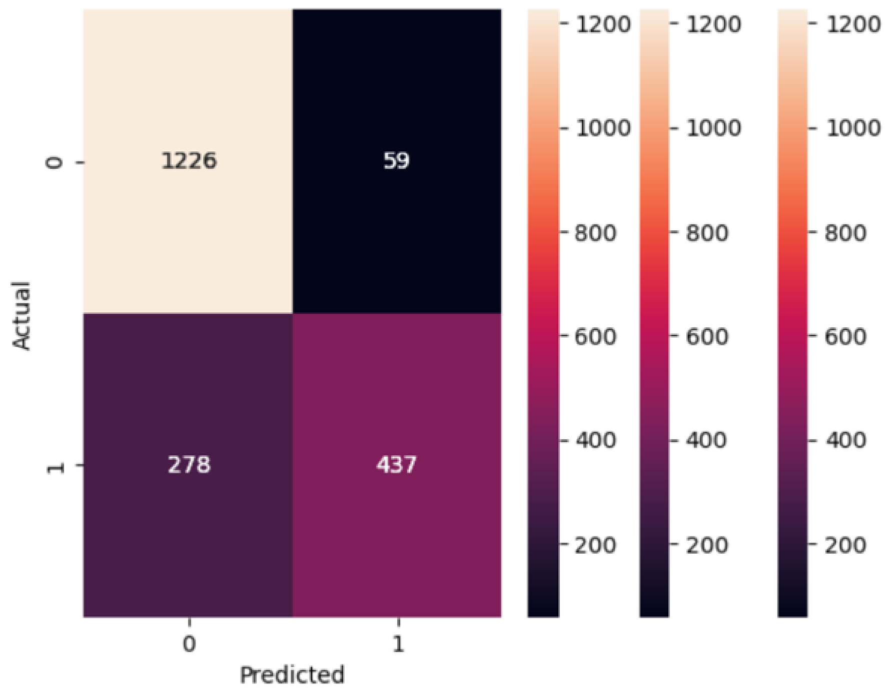

Figure 1. Decision tree accuracy—Confusion matrix.

The confusion matrix for DT1, DT2, and DT3 visualizes their actual versus predicted values, divided into four quadrants. The top left quadrant shows True Positives (TP), with 1226 cases accurately identified as positive. The top right quadrant shows False Positives (FP), with 59 cases incorrectly labeled as positive. The bottom left quadrant indicates False Negatives (FN), where 278 positive cases were misclassified as negative, and the bottom right quadrant contains True Negatives (TN), with 437 cases correctly identified as negative. DT1 and DT2 both exhibit strong predictive capabilities, achieving an accuracy of 83.15 percent, calculated by dividing the sum of true positives and negatives (1663) by the total predictions (2000).

4. Analysis

The analysis of DT models—DT1, DT2, and DT3—demonstrates a complex interaction between the models' parameters and their predictive outcomes. Changes in tree depth and node-splitting criteria significantly affect the accuracy of these models. Notably, DT2, optimized for depth and sample splits, exhibits a high recall rate, suggesting its strong ability to identify true positives, albeit at the cost of reduced precision, indicating a preference for sensitivity over specificity.

The evaluation of feature importance, as shown in **Tables 4** and **5**, highlights the critical variables that drive academic success predictions. Attendance, Age, and Pell Grant eligibility consistently emerge as the most influential factors across all three models, underscoring their pivotal roles in predicting graduation outcomes. For DT1 (**Figure 2**), Attendance is particularly significant, contributing to 39% of the model's focus, followed by Age and Pell Grant eligibility. This pattern is similarly observed in DT2 (**Figure 3**), with slight variations in the percentage contributions of these features.

Table 4. Decision tree model parameters and accuracy.

Model	Max_Depth	Min_Samples_Split	Accuracy
DT1	5	2	0.8315
DT2	3	3	0.8315
DT3	2	3	0.8085

Table 5. Decision tree feature importance.

Model	Attendance	Age	Pell	Entry GPA	Generation	Gender	Ethnicity
DT1	0.39	0.34	0.24	0.03	0.02	0.02	0.01
DT2	0.40	0.36	0.25	0.01	0.01	0.00	0.00
DT3	0.53	0.45	0.01	0.00	0.00	0.00	0.00

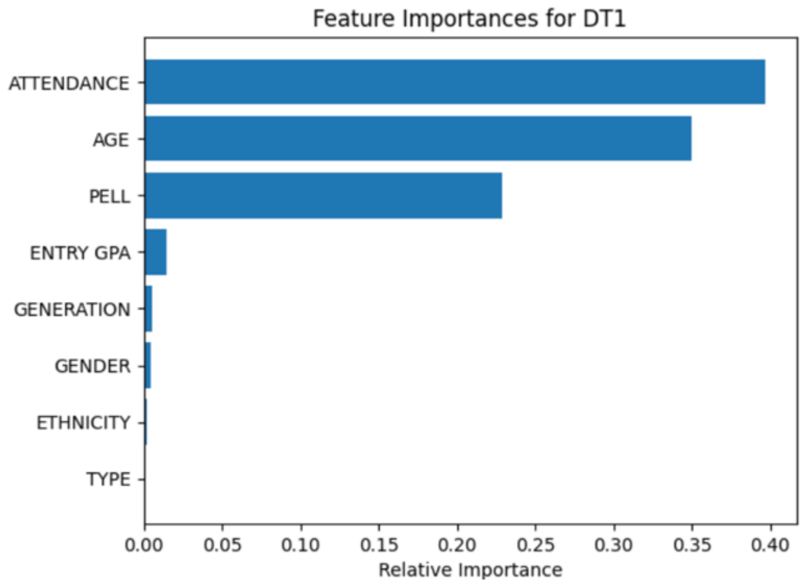

Figure 2. DT1 feature importance.

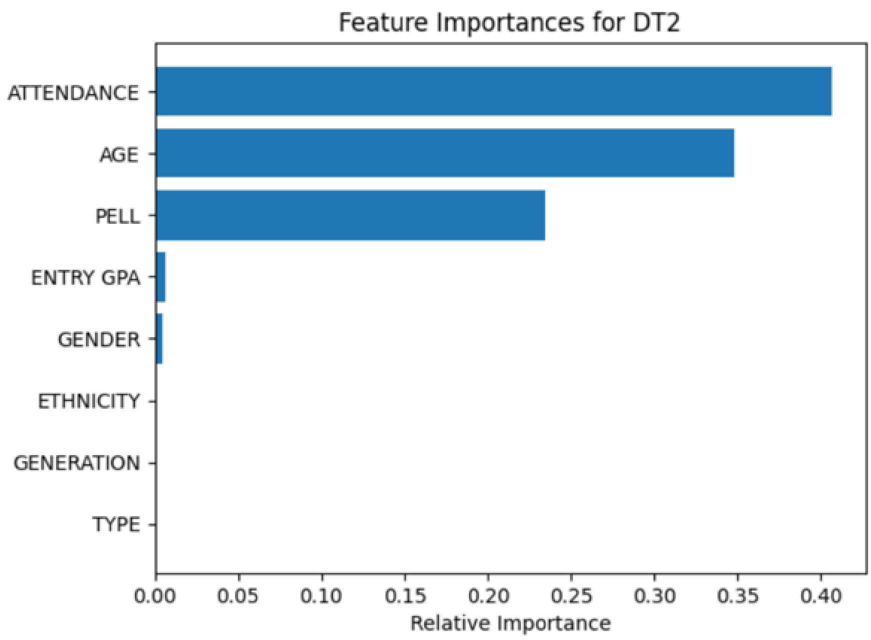

Figure 3. DT2 feature importance.

DT3 (**Figure 4**) shows a marked increase in the importance assigned to Attendance and Age, emphasizing the crucial role of student engagement and life stage in academic success. Interestingly, DT3 introduces Gender with a 5% influence, hinting at its potential, albeit minor, relevance in educational outcomes. These insights across the models underline the necessity for educational strategies that enhance student engagement, address financial barriers, and consider the diverse needs related to student age and, to some extent, gender, thereby guiding targeted interventions to boost graduation rates.

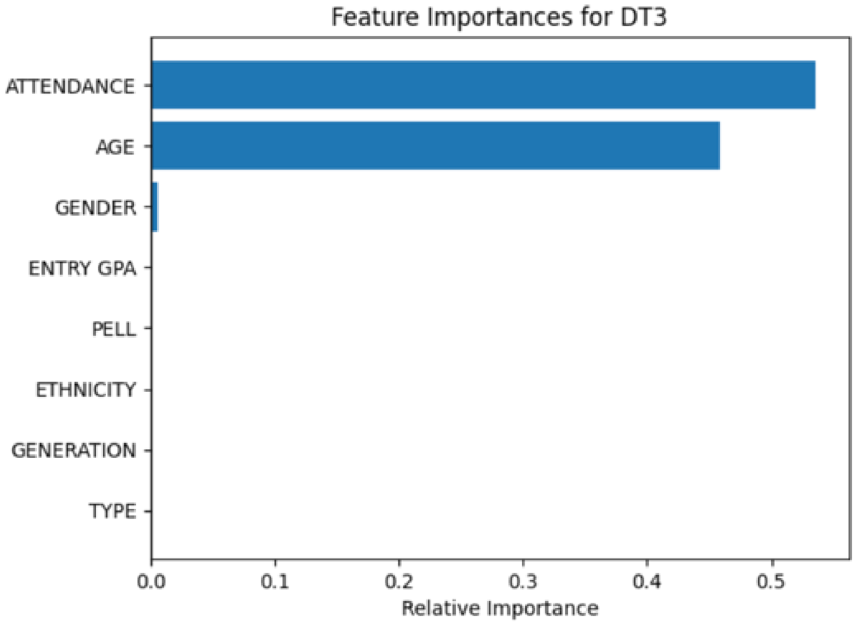

Figure 4. DT3 feature importance.

5. Discussion

The analysis of DT models provides valuable insights for educational institutions aiming to utilize predictive analytics to enhance student success and retention. These findings underscore several critical factors that influence academic outcomes for adult learners.

5.1. Importance of student engagement

The analyses highlight the critical importance of student engagement, particularly regular attendance and active participation, in predicting academic success. Policies and interventions designed to boost attendance and engagement could significantly enhance educational outcomes. This aligns with previous research that emphasizes the role of student engagement in academic achievement [3,5]. The feature importance analysis consistently showed Attendance as one of the top predictors across all DT models, reinforcing its vital role.

5.2. Effects of age and financial aid

Understanding the effects of age and financial aid on student performance can help institutions tailor support services more effectively to meet the needs of adult learners. The findings indicate that age and Pell Grant eligibility are significant predictors of academic success. This corroborates previous studies that have identified financial support as a crucial factor in promoting student retention and success [4,6]. The DT models revealed that older students and those receiving financial aid tended to have different educational trajectories, suggesting the need for targeted support interventions for these groups.

5.3. Gender as a predictive factor

The introduction of gender as a predictive factor, albeit with a lesser weight, indicates a complex interaction of identity factors in educational achievement. DT3's inclusion of Gender, while minor, suggests emerging insights into how demographic factors can influence educational pathways. This finding aligns with broader educational research that highlights the intersectionality of gender with other demographic factors affecting academic outcomes [7].

5.4. Model sensitivity and parameter adjustments

The sensitivity of DT models to adjustments in parameters such as Max_Depth and Min_Samples_Split is evident from the analysis. DT2, with its moderate depth, maintained high recall capabilities but showed variations in precision. This illustrates the nuanced balance between identifying true positives and minimizing false positives, a common challenge in predictive modeling. The performance of DT2 aligns with Gafarov et al. [45], who noted the importance of parameter tuning in optimizing model performance.

5.5. Feature importance

The feature importance analysis indicated a shift in emphasis towards Attendance and Age as primary predictors of academic success. DT3, in particular, increased focus on these aspects, highlighting their critical role in educational outcomes. This shift mirrors findings from Jacob and Henriques [43], who identified similar predictors in their educational data mining study. The consistent emphasis on these features across different models and studies underscores their significance in shaping educational strategies.

5.6. Implications for educational institutions

These results lay the groundwork for educational institutions to develop targeted strategies that cater to the diverse needs of their student populations. By focusing on enhancing engagement, addressing financial barriers, and recognizing the nuanced influences of demographic factors, institutions can create more inclusive and supportive educational environments. This approach is supported by recent studies that advocate for data-driven, personalized educational strategies to improve student outcomes [33,44]. The findings from the DT analyses, therefore, provide a robust foundation for educational institutions to implement predictive analytics in their strategies for student success. The insights gained from this study contribute to the growing body of research on the use of machine learning in higher education, highlighting both the potential and the challenges of these advanced analytical tools.

5.7. Limitations and future research

This study, while comprehensive, has several limitations that must be acknowledged to provide context for interpreting the findings and guiding future research. Firstly, the dataset used in this study spans from the academic years 2013–2014 to 2021–2022. Although this provides a broad temporal scope, it may not capture the most recent trends and developments in higher education, particularly those influenced by the COVID-19 pandemic. Additionally, the geographical constraints of the dataset, being limited to a specific region, may affect the generalizability of the findings to other regions with different educational systems and socioeconomic conditions. Furthermore, while key variables such as age, attendance, and Pell Grant eligibility were included, other potentially influential factors like part-time versus full-time enrollment, online versus in-person learning, and specific support services utilized were not considered.

Secondly, there are methodological constraints inherent in the study. The primary utilization of DT models, while effective, has limitations in handling certain types of data and interactions between variables. Other ML models, such as NN or ensemble methods, might better capture these complexities. Additionally, the preprocessing steps involved standard techniques, but more advanced feature engineering could enhance model performance and uncover deeper insights.

Thirdly, limitations regarding sample size and diversity must be noted. The final dataset, refined to 9999 records, may not fully represent the broader adult learner population, particularly concerning diversity in demographic characteristics and educational backgrounds. Moreover, there is a potential for sampling bias, as the dataset might overrepresent certain groups of students, such as those more likely to apply for Pell Grants, while underrepresenting others.

Future research can build on this study by exploring how age, attendance, and Pell Grant eligibility affect adult student success through several avenues. Firstly, qualitative studies could provide valuable insights. A qualitative study could explore the influence of learning experiences, study habits, and cultural backgrounds across different adult age groups by interviewing adults who have returned to higher education. This approach would help capture deeper insights into the qualitative aspects that quantitative data alone cannot provide.

Secondly, revisiting the role of GPA as a predictor of success presents another important research avenue. Longitudinal studies could track changes in GPA's predictive power over multiple academic periods and life stages. Additionally, research could examine the specific impact of GPA on the initial and subsequent college attempts of adult learners, comparing how GPA's influence may change as students return to school after significant life experiences.

Thirdly, broadening research to include factors like socioeconomic status, employment, and family responsibilities would offer a comprehensive view of the student experience through both quantitative and qualitative analyses. This broader scope could provide a more nuanced understanding of the challenges and supports needed by adult learners. Moreover, investigating faculty and policymaker perspectives on machine learning and data-driven decision-making could reveal how these tools shape institutional policies, particularly when data challenges traditional views. This research could inform resource allocation and institutional changes, ensuring that policies are both data-driven and responsive to the needs of adult learners. There is also potential to explore how higher education institutions manage technological advancements and the implementation of ML, focusing on data collection and operational capabilities. Thus, these research directions highlight the importance of a nuanced, data-driven approach in making educational decisions that could make higher education more inclusive, adaptable, and supportive for adult learners. The findings from such research could significantly influence policy formulation, refine institutional practices, and enhance services for adult students.

6. Conclusion

In the evolving landscape of higher education, the application of machine learning (ML) techniques, particularly Decision Tree (DT) models, has garnered significant attention for their potential to address the unique challenges faced by adult learners. This study sought to explore the efficacy of Classification and Regression Trees (CART) in predicting academic success among adult learners, focusing on key variables such as age, attendance, Pell Grant eligibility, and other demographic factors. Adult learners represent a substantial and growing segment of the student population in the United States. They encounter a myriad of socio-cultural, economic, and institutional obstacles that impact their educational experiences and outcomes. Despite their increasing presence, there is a notable gap in tailored educational strategies that address the specific challenges of adult learners. This study aimed to fill this gap by leveraging ML techniques to provide data-driven insights that can inform the development of inclusive and effective educational policies and practices.

The study utilized a comprehensive dataset spanning from the academic years 2013–2014 to 2021–2022, encompassing variables such as age, ethnicity, gender, Pell Grant eligibility, and academic performance metrics. A quantitative evaluation of CART DT models was conducted to assess their effectiveness in predicting degree completion rates among adult learners. Key statistical measures such as accuracy, precision, recall, and F1 score were used to evaluate model performance. The dataset was divided into training and testing subsets using an 80/20 split ratio to ensure a balanced allocation for model calibration and validation.

The analysis revealed several critical insights. Attendance emerged as a primary predictor of academic success, underscoring the importance of student engagement. Age and Pell Grant eligibility were also significant predictors, indicating the need for tailored support services for older students and those receiving financial aid. Gender, while a minor factor, provided emerging insights into the complex interaction of demographic variables in educational outcomes. The sensitivity of DT models to parameter adjustments highlighted the necessity of fine-tuning to optimize predictive accuracy. These findings align with and extend previous research, contributing to a deeper understanding of the factors influencing adult learner success.

While this study provides valuable insights, several avenues for future research are recommended:

Qualitative Studies: Conduct qualitative studies to explore the nuanced experiences of adult learners, focusing on learning experiences, study habits, and cultural backgrounds across different age groups.

Longitudinal Analysis of GPA: Revisit the role of GPA as a predictor of success by applying new data in longitudinal studies to track its predictive power over time and across different life stages.

Broader Socioeconomic Factors: Expand research to include a wider range of socioeconomic factors, such as employment status and family responsibilities, to provide a comprehensive view of the adult learner experience.

Faculty and Policymaker Perspectives: Investigate the perspectives of faculty and policymakers on the use of ML and data-driven decision-making in higher education, focusing on how these tools can shape institutional policies and practices.

Technological Integration: Explore how higher education institutions manage the integration of ML technologies, with a focus on data collection practices and operational capabilities.

By addressing these areas, future research can further enhance the understanding and support of adult learners, contributing to more equitable and effective educational environments.

This study underscores the potential of ML techniques, particularly DT models, to provide actionable insights that can transform educational strategies for adult learners. The findings highlight the importance of data-driven approaches in developing tailored interventions that enhance student engagement, address financial barriers, and recognize the diverse needs of the student population. As higher education continues to evolve, these insights will be crucial in creating inclusive and supportive learning environments that foster the success of all students.

Author contributions: Conceptualization, EB; methodology, EB and KP; validation, EB and KP; investigation, EB; original draft preparation, JH; writing—review and editing, JH; visualization, EB. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

Abbreviations

CART	Classification and Regression Trees
ML	Machine Learning
RF	Random Forests
DT	Decision Tree
ID3	Iterative Dichotomiser 3
ANNs	Artificial Neural Networks
XGBoost	eXtreme Gradient Boosting
NN	Neural Networks
SMS	student management system
GBM	Gradient Boosting Machine
CV	cross-validation
TP	True Positives
FP	False Positives
FN	False Negatives
TN	True Negatives
LMT	Logistic Model Tree

References

1. Job MA, Pandey J. Academic Performance Analysis Framework for Higher Education by Applying Data Mining Techniques. In: Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO); 2020. doi: 10.1109/icrito48877.2020.9197925
2. Caspari-Sadeghi S. Learning assessment in the age of big data: Learning analytics in higher education. Cogent Education. 2022; 10(1). doi: 10.1080/2331186x.2022.2162697

3. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist. 2000; 55(1): 68–78. doi: 10.1037/0003-066x.55.1.68
4. Nieto Y, García-Díaz V, Montenegro C, et al. Supporting academic decision making at higher educational institutions using machine learning-based algorithms. Soft Computing. 2018; 23(12): 4145–4153. doi: 10.1007/s00500-018-3064-6
5. Musso MF, Hernández CFR, Cascallar EC. Predicting key educational outcomes in academic trajectories: A machine-learning approach. Higher Education. 2020; 80(5): 875–894. doi: 10.1007/s10734-020-00520-7
6. Teng Y, Zhang J, Sun T. Data-driven decision-making model based on artificial intelligence in higher education system of colleges and universities. Expert Systems. 2022; 40(4). doi: 10.1111/exsy.12820
7. Alkhalil A, Abdallah MAE, Alogali A, et al. Applying Big Data Analytics in Higher Education. International Journal of Information and Communication Technology Education. 2021; 17(3): 29–51. doi: 10.4018/ijicte.20210701.oa3
8. Chaurasia SS, Kodwani D, Lachhwani H, et al. Big data academic and learning analytics. International Journal of Educational Management. 2018; 32(6): 1099–1117. doi: 10.1108/ijem-08-2017-0199
9. National Student Clearinghouse Research Center. Some college, no credential student Outcomes: Annual progress report—academic year—2021/22. Available online: https://nscresearchcenter.org/some-college-no-credential/ (accessed on 1 June 2024).
10. Sadiq MH, Ahmed NS. Classifying and Predicting Students’ Performance using Improved Decision Tree C4.5 in Higher Education Institutes. Journal of Computer Science. 2019; 15(9): 1291–1306. doi: 10.3844/jcssp.2019.1291.1306
11. Yağcı M. Educational data mining: prediction of students’ academic performance using machine learning algorithms. Smart Learning Environments. 2022; 9(1). doi: 10.1186/s40561-022-00192-z
12. Hamoud AK, Hashim AS, Awadh WA. Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. International Journal of Interactive Multimedia and Artificial Intelligence. 2018; 5(2): 26. doi: 10.9781/ijimai.2018.02.004
13. Marzuqi A, Laksitowening KA, Asror I. Temporal Prediction on Students’ Graduation using Naïve Bayes and K-Nearest Neighbor Algorithm. Jurnal Media Informatika Budidarma. 2021; 5(2): 682. doi: 10.30865/mib.v5i2.2919
14. Widaningsih S, Muhamad W, Hendriyanto R, et al. An ID3 Decision Tree Algorithm-Based Model for Predicting Student Performance Using Comprehensive Student Selection Data at Telkom University. Ingénierie des systèmes d information. 2023; 28(5): 1205–1212. doi: 10.18280/isi.280508
15. Quinlan JR. Induction of decision trees. Machine Learning. 1986; 1(1): 81–106. doi: 10.1007/bf00116251
16. Gotardo MA. Using Decision Tree Algorithm to Predict Student Performance. Indian Journal of Science and Technology. 2019; 12(8): 1–8. doi: 10.17485/ijst/2019/v12i5/140987
17. Alsariera YA, Baashar Y, Alkawsi G, et al. Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. Computational Intelligence and Neuroscience. 2022; 2022: 1–11. doi: 10.1155/2022/4151487
18. Martins M, Miguéis V, Fonseca D, Alves A. (2019). A data mining approach for predicting academic success: A case study. In: Rocha AC. Ferrás M. (editors). Information Technology and Systems: Advances in intelligent systems and computing. Springer; 2019. Volume 918. pp. 45–56. https://doi.org/10.1007/978-3-030-11890-7_5
19. Garg A, Lilhore UK, Ghosh P, et al. Machine Learning-based Model for Prediction of Student’s Performance in Higher Education. In: Proceedings of the 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN); 2021. doi: 10.1109/spin52536.2021.9565999
20. Ahmad Tarmizi, S. S., Mutalib, S., Abdul Hamid, N. H., Abdul-Rahman, S., & Md Ab Malik, A. (2019). A case study on student attrition prediction in higher education using data mining techniques. In Soft Computing in Data Science: 5th International Conference, SCDS 2019, Iizuka, Japan, August 28–29, 2019, Proceedings 5 (pp. 181-192). Springer Singapore.
21. Ivanov A. Decision Trees for Evaluation of Mathematical Competencies in the Higher Education: A Case Study. Mathematics. 2020; 8(5): 748. doi: 10.3390/math8050748
22. Crismayella Y, Satyahadewi N, Perdana H. Algoritma Adaboost pada Metode Decision Tree untuk Klasifikasi Kelulusan Mahasiswa. Jambura Journal of Mathematics. 2023; 5(2): 278–288. doi: 10.34312/jjom.v5i2.18790
23. Nguyen MK, Huynh VT, Nguyen HD. Predicting graduation grades using Machine Learning: A case study of Can Tho University students. CTU Journal of Innovation and Sustainable Development. 2023; 15(ISDS): 83–92. doi: 10.22144/ctujoisd.2023.038
24. Акмеше ОФ, Кьор Х, Ербей Х. Use of machine learning techniques for the forecast of student achievement in higher

education. Information Technologies and Learning Tools. 2021; 82(2): 297–311. doi: 10.33407/itlt.v82i2.4178
25. Zhou J, Ye J min. Sentiment analysis in education research: A review of journal publications. Interactive Learning Environments. 2020; 31(3): 1252–1264. doi: 10.1080/10494820.2020.1826985
26. Oreški D, & Zamuda D. Machine Learning Based Model for Predicting Student Outcomes. IEOM Society International. 2022. doi: 10.46254/an12.20220967
27. Khor ET. A data mining approach using machine learning algorithms for early detection of low-performing students. The International Journal of Information and Learning Technology. 2022; 39(2): 122–132. doi: 10.1108/ijilt-09-2021-0144
28. Yan J, Chi X. Analysis and Prediction of College Students' Employment based on Decision Tree Classification Algorithm. In: Proceedings of the 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE); 2023. doi: 10.1109/icdcece57866.2023.10150841
29. Meng Z. Analysis and Prediction of College Students' Employment Based on Decision Tree Classification Algorithm. In: Proceedings of the 2023 World Conference on Communication & Computing (WCONF); 2023. doi: 10.1109/wconf58270.2023.10234985
30. Pradana MG, Rangga Pinastawa IW, Maulana N, et al. Performance Analysis of Tree-Based Algorithms in Predicting Employee Attrition. CCIT Journal. 2023; 16(2): 220–232. doi: 10.33050/ccit.v16i2.2580
31. Yin L. The Application of Improved Decision Tree Algorithm in College Employment Management. In: Proceedings of the 2023 3rd Asian Conference on Innovation in Technology (ASIANCON); 2023. doi: 10.1109/asiancon58793.2023.10270643
32. Muzumdar P, Basyal GP, Vyas P. An Empirical Comparison of Machine Learning Models for Student's Mental Health Illness Assessment. Asian Journal of Computer and Information Systems. 2022; 10(1). doi: 10.24203/ajcis.v10i1.6882
33. Battista K, Diao L, Patte KA, et al. Examining the use of decision trees in population health surveillance research: An application to youth mental health survey data in the COMPASS study. Health Promotion and Chronic Disease Prevention in Canada. 2023; 43(2): 73–86. doi: 10.24095/hpcdp.43.2.03
34. Li H. Analysis of mental health influencing factors and students' psychological education countermeasures based on ID3 algorithm. Applied Mathematics and Nonlinear Sciences. 2023; 9(1). doi: 10.2478/amns.2023.2.01377
35. Xiaocheng H. Application of Decision Tree Algorithm in College Students' Mental Health Evaluation. In: Proceedings of the 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS); 2023. doi: 10.1109/icicacs57338.2023.10099654
36. Krishnan R, Kumari S, Badi AA, et al. Predictive machine learning model for mental health issues in higher education students due to COVID-19 using HADS assessment. Arab Gulf Journal of Scientific Research. 2023. doi: 10.1108/agjsr-01-2023-0034
37. Rueda E, Lowe Swift C. Academic Belonging in Higher Education. Routledge; 2023. doi: 10.4324/9781003443735
38. Salihoun M. State of Art of Data Mining and Learning Analytics Tools in Higher Education. International Journal of Emerging Technologies in Learning (IJET). 2020; 15(21): 58. doi: 10.3991/ijet.v15i21.16435
39. Tahiru F, Parbanath S, Agbesi S. Machine Learning-based Predictive Systems in Higher Education: A Bibliometric Analysis. Journal of Scientometric Research. 2023; 12(2): 436–447. doi: 10.5530/jscires.12.2.040
40. Mao H, Fang Z, Xia Z. Model of the Influence of Students' Participation on Learning Effects Using Educational Artificial Intelligence. In: Proceedings of the 2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA). 2023. doi: 10.1109/icccbda56900.2023.10154663
41. Pinto AS, Abreu A, Costa E, et al. How Machine Learning (ML) is Transforming Higher Education: A Systematic Literature Review. Journal of Information Systems Engineering and Management. 2023; 8(2). doi: 10.55267/iadt.07.13227
42. Feng L. Research on Higher Education Evaluation and Decision-Making Based on Data Mining. Scientific Programming. 2021; 2021: 1–9. doi: 10.1155/2021/6195067
43. Jacob D, Henriques R. Educational Data Mining to Predict Bachelors Students' Success. Emerging Science Journal. 2023; 7: 159–171. doi: 10.28991/esj-2023-sied2-013
44. Attiya WM, Shams MB. Predicting Student Retention in Higher Education Using Data Mining Techniques: A Literature Review. In: Proceedings of the 2023 International Conference On Cyber Management and Engineering (CyMaEn); 2023. doi: 10.1109/cymaen57228.2023.10051056
45. Gafarov, M. F., Okishev, K. Y., Makovetskiy, A. N., Pavlova, K. P., & Gafarova, E. A. (2023). Construction of Models for Predicting the Microstructure of Steels after Heat Treatment Using Machine Learning Methods. Steel in Translation, 53(11), 1120-1129.

46. Chen R. Research on the development path of higher education model innovation based on quadratic planning algorithm. Applied Mathematics and Nonlinear Sciences. 2023; 9(1). doi: 10.2478/amns.2023.2.01100