EPiC Series in Built Environment
Volume 7, 2026, Pages 41–50
Proceedings of Associated Schools of Construction 62nd Annual International Conference

Artificial Intelligence Decision Tree Model of Construction Students' Four-Year Outcome

Tulio Sulbaran[1], Jared Burgoon,[2] Olugbenro Ogunrinde[3], Rachel Mosier[4]
[1]University of Texas at San Antonio, [2]Kansas State University, [3]Tarleton State University, [4]Oklahoma State University

The construction industry faces a critical shortage of skilled professionals, making student persistence and on-time graduation within construction management programs a vital concern. The objective of this research was to identify and map critical student characteristics associated with four-year graduation within a university-level construction workforce pathway program. The study employed a quantitative research methodology to analyze fourteen years of institutional research data from n=613 undergraduate student records. Specifically, a transparent and interpretable Artificial Intelligence (AI) Decision Tree model was utilized to analyze this extensive dataset and classify it into the likelihood of graduating, or not, within four years based on factors influencing the students. The results show a significant disparity in students graduating within four years based on certain student characteristics. Similarly, there are other characteristics that make the students less likely to graduate. This second finding, in the authors' opinion, is the most important as it allows universities to use their limited resources to assist these students in increasing their graduation rate to join the construction industry with their degrees. The intellectual merit of this research lies in its enhanced understanding of students' characteristics who do not graduate within four years. This study's broader impact is that it provides data-driven, actionable insights for universities to implement targeted interventions, potentially mitigating the critical shortage of skilled construction professionals.

Keywords: Artificial Intelligence, Construction, Graduation, Intervention, Persistence, Students

Introduction

The construction industry currently faces a critical and growing shortage of skilled professionals, a challenge that threatens long-term industry growth and project capacity. This shortage of skilled professionals has a significant direct impact on cost performance (Karimi et al., 2018). Therefore, increasing the number of university construction graduates is important. While higher education institutions need to improve student recruitment (J. M. Burgoon et al., 2024) into Construction Management (CM) programs, a potentially more impactful and resource-efficient strategy is to enhance student persistence and four-year graduation rates within existing university construction workforce pathway programs. Degree completion, time to degree, and graduation rates are of utmost importance in academia generally, (Müller et al., 2023) and should be for the construction industry in order to supply the necessary talent to the workforce pipeline. Despite the concerning statistics, there is limited published research that addresses why students struggle and fail to complete their degree within four years, or even abandon their education (Müller et al., 2023). Therefore, data-driven analysis is needed to better understand the factors and characteristics that stratify and predict construction students' graduation outcomes within four years.

For many years, predictive modeling has been used to analyze input data to produce the best possible predictions of an outcome variable (Cranmer & Desmarais, 2017). However, in the last few years, explainable applications of Artificial Intelligence (AI) machine learning (ML) have gained significant attention from both academics and practitioners due to their ability to handle large datasets, uncover complex relationships, and improve predictive accuracy (Alanazi, 2025). Thus, the objective of this study was to identify and map critical student characteristics to determine their association with four-year graduation outcomes, using an interpretable Artificial Intelligence (AI) Decision Tree model to predict the outcome of graduation (Y) or non-graduation (N) within four years.

Literature Review

Predicting student performance is important in the educational domain because student status analysis helps improve the performance of higher education institutions (Salah Hashim et al., 2020). Student persistence and timely graduation are critical success indicators for both higher education institutions and the industries they serve. Unfortunately, despite the fact that the problem of dropout in higher education has been studied for decades (Castro-Montoya et al., 2025), the problem persists today. This is particularly important for the construction sector, where there is a direct linkage between program completion rates and the availability of qualified construction managers. This linkage underscores the importance of having construction management education programs that can produce qualified graduates with the requisite skills, with timely graduation to meet the growing needs of the construction industry (Whitmore & Ojajuni, 2023). Particularly, in the field of Construction Management, student program retention is further complicated by the abundant work opportunities, rigorous technical coursework, and the demand for field-specific experience prior to graduation.

Student background characteristics are widely recognized as strong predictors of success at higher education institutions. This research focuses on three established high-impact factors:

1. <u>Transfer Student Status:</u> This group is defined as those who begin or attend an institution (e.g., two-year or four-year college) and transfer to another institution (e.g., four-year university) (Collins, 2021). Some of the challenges that transfer students face include social networks, peer relationships, sense of belonging, motivation to study, and academic workload stress, among others (Lau et al., 2022)

2. <u>Pell Grant Recipient Status:</u> Pell Grants are awarded to students from low-income families (Lau et al., 2022). Therefore, the receipt of a Pell Grant serves as a robust proxy for socioeconomic disadvantage and low-income background. Students from low-income families frequently face financial stress due to increased obligations to work outside of school, which limits the time available for study and campus involvement. Thus, some studies suggest that Pell recipients face more challenges to graduate (Lau et al., 2022) in general.

3. <u>First-Generation Status:</u> These are students who do not have a parent who obtained a four-year degree. Research indicates these students often have higher work and family demands than their peers, have lower academic preparation, and are less likely than their peers to graduate (J. Burgoon & Elliott, 2021).

Machine learning (ML) has emerged as a transformative tool for strategic business decision-making and sustainable competitive advantage (Alanazi, 2025; Olayinka, 2019). ML has numerous advantages over traditional methods in decision-making, as it allows real-time processing of vast datasets, can incorporate a variety of factors simultaneously, and predict future trends more accurately than traditional methods (Alanazi, 2025). One ML model is a decision tree, which is among the best studied and most widely used ML tools (Blockeel et al., 2023). An AI Decision Tree is a structure whose internal nodes can test input data patterns and whose leaf nodes can be taken as categories or results of those input patterns (Navada et al., 2011). Unlike complex "black-box" models (such as deep neural networks), Decision Trees provide a clear, hierarchical set of "if-then-else" rules. AI Decision Trees are flowchart-like structures that model decisions and their possible consequences (Jimenez-Roa et al., 2023). This transparent flowchart-like structure is crucial for university administrators, as the model's structure can be directly translated into practical, targeted interventions. AI Decision Trees are used by decision makers, mainly due to their intuitive interpretation (Jimenez-Roa et al., 2023). By identifying the sequential splits that lead to high-risk leaf nodes, AI Decision Tree models allow institutions to focus their limited resources on the specific student subgroups that are most vulnerable to attrition.

Methodology

The study employed a quantitative research methodology, utilizing historical institutional student records data obtained from the Institutional Research office of the institution. The dataset was organized in a spreadsheet format where each row represented an individual student's record, and each column corresponded to an individual variable. This structure was used because it enabled direct input into the supervised machine learning approach centered on an Artificial Intelligence (AI) Decision Tree model. This model was chosen due to its ability to perform classification, predicting a critical binary outcome, such as student graduation or non-graduation. Furthermore, the AI Decision Tree results are interpretable and easy to grasp, and have stood the test of time (Costa & Pedreira, 2023). The research methodology included the following four steps:

Step 1- Problem Definition

As shown in Table 1, the dependent or target variable (what to be predicted) was identified as corresponding to whether or not a student will graduate within four years (Dependent Variable). The independent or input/feature variables (what will be used for prediction) were also determined and included: (1) Pell Recipient, (2) Transfer Students, and (3) First Generation Student, which are commonly available to program leaders through institutional student records.

Table 1. Model Variables

Target Variable (Dependent Variable)	**Input/Feature Variables (Independent Variable)**
Student graduation within four years (Y/N)	Pell Recipient (Y/N)
	Transfer Student (Y/N)
	First Generation (Y/N)

Step 2- Model Training and Representation

The data collected was split into training (~80%) and testing (~20%) sets. This data slip was selected because empirical analysis has shown that the best results are attained (Gholamy et al., 2018). The training dataset was used to let the AI Decision Tree learn and map the complex relationship between the independent / input variables and the dependent / target variable, generating transparent "if-then" decision rules that provide clear, actionable insights into the factors influencing student success.

The AI model was presented as a tree with multiple levels. At each level, each node in the tree represents a decision (split) based on one of the values of the input/feature variables. The last row (bottom) boxes of the decision tree are leaf nodes that represent the final prediction corresponding to whether or not the student will graduate within four years.

Each node box included five pieces of information:

1. Decision Condition: indicates the variable being evaluated and the evaluation criteria. If the evaluation criterion is True, the path goes left, and if not (False), the path goes right.
2. Gini: corresponds to the Gini impurity index, a measure of how mixed the classes are in that node. 0.0 = pure node (all one class) and 0.5 = perfectly mixed (50% Y, 50% N)
3. Samples: percentage of total data in that node.
4. Value = Proportion of each class in that node corresponding to the predicted outcome, with the left value corresponding to "No", the student will not graduate, and the right side corresponding to "Yes", the student will graduate.
5. Class = Predicted class for that node (the majority class). In other words, it will be "No" if more than fifty percent (>50%) of the students in this node will not graduate.

Step 3 - Model Evaluation

The AI Decision tree's performance was evaluated using the unseen testing dataset. Evaluating using the unseen testing dataset was important to ensure the model's ability to generalize its predictions to new, real-world data, rather than just having memorized the training data (overfitting). The evaluation metrics included: Accuracy, Precision, Recall, and F1-Score. These evaluation metrics quantify different aspects of the model's predictive capability based on the relationship between True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

1. Accuracy: The overall proportion of correctly classified instances out of the total number of instances evaluated (Sulbaran, 2023). It provides a general measure of how often the model is right.

$$Accuracy = \frac{TP + TN}{TP + FP + FB + TN}$$

2. Precision: represents the proportion of positive samples that were correctly classified to the total number of positive predicted samples, as indicated (Tharwat, 2020)

$$Precision = \frac{TP}{TP + FP}$$

3. Recall: Measures the completeness of the positive predictions (Sulbaran, 2023). In other words, of all the cases that were actually positive (e.g., students who did graduate), how many did the model correctly identify? $Recall = \frac{TP}{TP+FN}$

4. F1-Score: represents the harmonic mean of precision and recall (Tharwat, 2020). Combines Precision and Recall metrics into a single score, providing a balanced measure of the model's performance.

$$F1 - Score = \frac{Precision\ x\ Recall}{Precision + Recall}$$

Step 4 – Model Results Interpretation and Discussion

The interpretation of the model results was conducted through a direct analysis of the Decision Tree results visualization. Focusing on the class distribution (represented by the value = [N, Y] of graduation). The mode was traced from the top node passing through each level to the last level (leaf nodes) that corresponded to the predicted outcome N or Y to graduate. This resulted in a specific demographic cohort with a rating from 1 to 8 representing the highest and lowest likelihood of graduating in 4 and 6 years. This was accomplished by quantitatively determining the rating based on the highest proportion of the positive class ('Y') in the value array of the leaf.

Results

Demographic Information

The study included 613 undergraduate student records from 2004 to 2018. The year 2004 was selected because that is the oldest data available, and the 2018 data was selected to ensure that by the time of the study, the students had been at the university for at least 6 years. There was no missing data as all characteristics were required field in the system The records indicated that 48.1% (n = 295) of the students transferred from another institution, and 51.9% (n = 318) were first-time, non-transfer students. Less than half of the students, 47.5% (n = 291), were Pell Grant recipients, indicating that they came from low-income backgrounds, whereas 52.5% (n = 322) did not receive Pell support. The analysis revealed that 52.9% (n = 324) of students were identified as first-generation college students, meaning they do not have a parent who graduated from college, while 47.1% (n = 289) had at least one parent with college experience. Only 20.6% (n = 126) of the students graduated in 4 years.

Resulting AI Decision Tree Model Representation

The AI Decision Tree Model used to predict the four-year graduate rate of construction students was trained with 79.9% of the dataset (n=490) and evaluated with the remaining 20.1% (n=123). The AI Decision Tree model is shown in Figure 1, classifying students into two possible outcomes (Classes): Y (Graduated) and N (Did not Graduate). The decision criteria within each node dictate the flow, with the "True" path to the left and the "False" path to the right. The resulting ranking of the likelihood of students' graduation is shown in the leaf nodes and augmented with numbered circles (1–8), where 1 represents the students most likely to graduate within four years and 8 are less likely to graduate.

At the first/top level of the tree (Root node), the AI model identified transfer status as the most significant factor. Transfer status is an independent binary variable where 0 means that the student is "No Transfer (Traditional Freshman Entry)" and 1 means the student is a "Transfer Student." Therefore, since the decision condition is "less than or equal to 0.5," the "No Transfer" students follow the left (True) branch, while values greater than 0.5 correspond to "Transfer" students on the right (False) branch.

This first decision is the primary split that encompasses 100% of the data (all students). At this root node, the model identified that 79.2% of students are not expected to graduate within four years, while the remaining 20.8% are expected to graduate. This node's Gini value of 0.33 indicates moderate impurity, suggesting that while "No Graduation" is dominant, there is still a portion of students who do graduate.

On the second level, the left branch represents No Transfer students, which account for 47.3% of the total sample. Within this group, 59.5% are not expected to graduate, while 40.5% are expected to graduate within four years. On the right branch of the second level, the Transfer students represent 52.7% of the total sample. Within this group, the Gini index is much lower (0.06), indicating a more homogeneous outcome, with most transfer students (approximately 96.9%) predicted not to graduate within four years, while only 3.1% are predicted to graduate. This pattern demonstrates that transfer students have a substantially lower four-year graduation likelihood than non-transfer students.

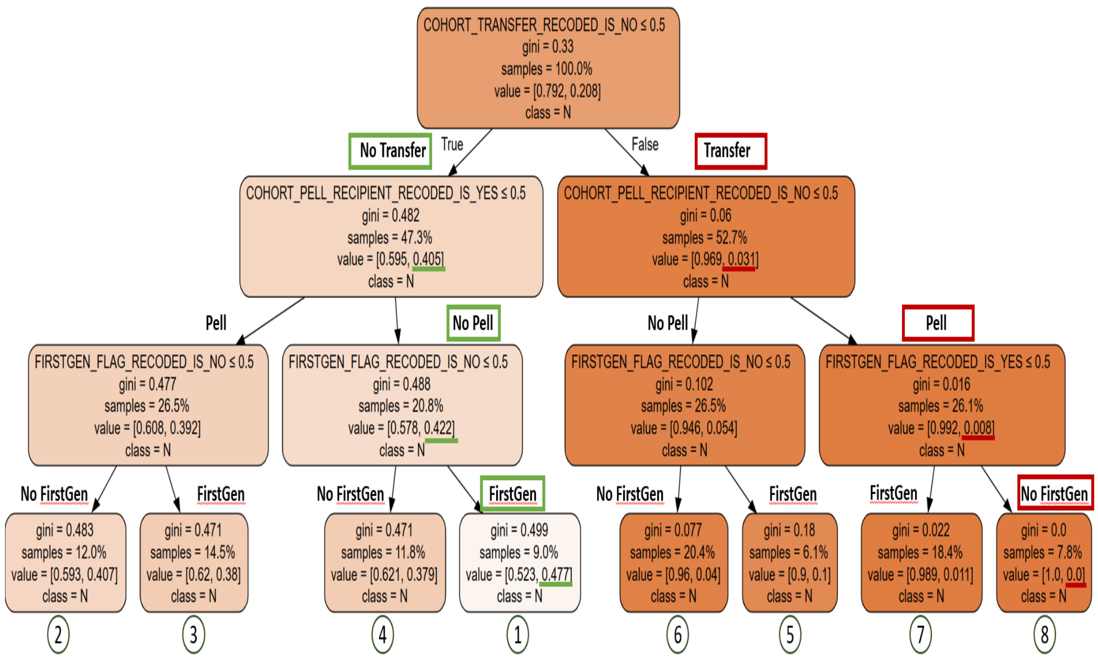

Figure 1. AI Decision Tree Model of Construction Students' Four-Year Graduate Rate

The third and fourth level branches show how Pell recipient status and first-generation status further refine the prediction. For instance, Transfer students who receive Pell Grants (third level rightmost branch) have a very low Gini value (0.016), indicating almost complete purity, meaning nearly all students in this subgroup are predicted not to graduate within four years. Conversely, No Transfer students who are not Pell recipients and first-generation (fourth middle branches) exhibit a 52.3%, which is a much higher predicted likelihood of graduating within 4 years, suggesting that financial stability and parental education are strong positive factors in graduation likelihood.

Figure 1 shows that graduation likelihood varies notably across student groups. The numbered circles at the bottom demonstrate that non-transfer students consistently show stronger graduation outcomes than transfer students, indicating that students who begin and remain at the same institution are more likely to complete their degrees within four years. Similarly, it can be observed at the third level of the tree that Pell recipients have a lower graduation likelihood with 39.2% (Non Transfer) and 0.8% (Transfer), while the Non-Pell recipients with 42.2% (Non Transfer) and 5.4% (Transfer), suggesting that having better financial stability contributes to a higher graduation rate.

Figure 1 illustrates how educational attainment outcomes are stratified by student background. Non-transfer and non–Pell recipient students have the highest likelihood of graduating within six years, while transfer and Pell recipients remain at greater risk of delayed graduation or attrition. These findings highlight the need for targeted institutional interventions such as transfer transition programs, mentorship initiatives, and financial aid support to promote persistence and degree attainment among the most vulnerable student populations, perhaps following Tinto's student integration model (Castro-Montoya et al., 2025)

The model was tested using 20.1% (n=123) of the data set. The model achieved an overall accuracy of 80.49%, meaning it correctly classified approximately four out of every five students in the test set. While this figure appears strong at first glance, a deeper analysis of the precision, recall, and F1-scores is important, particularly where the 'N' testing class (n=99) heavily outweighed the 'Y' class (n=24).

The model exhibited strong, reliable performance on the majority class, 'N' (Did not Graduate). With a Recall of 1.00, the model correctly identified 100% of the actual non-graduates (99 out of 99). Furthermore, a Precision of 0.80 indicates that when the model predicted a student would not graduate, it was correct 80.0% of the time. This resulted in a high F1-Score of 0.89, which indicates the model is highly effective at identifying construction students who will not graduate in four years.

Conversely, the model was not able to predict the minority class, 'Y' (Graduate). Both the Precision and Recall are 0.00 for the 'Y' class. This means the model failed to correctly identify any of the 24 students who actually graduated, and every single time the model predicted a 'Y' outcome, that prediction was incorrect. The resulting F1-Score of 0.00 confirms the model has no predictive power for graduates.

This outcome is a classic sign of Class Imbalance Bias or Overfitting to the Negative Class. The high 80.49% overall accuracy is achieved simply by correctly predicting the majority 'N' class, while the model completely ignores the minority 'Y' class. Due to the rarity of the 'Y' class in the training data, the model learned that the safest strategy is to predict 'N' (Not Graduate) for every single sample. This makes the Decision Tree effectively a constant predictor, which is only accurate because the 'N' class is the overwhelming majority. In short, despite the model's achieving an excellent overall accuracy (80.49%, the model's performance revealed a significant disparity in its ability to predict the two classes Y (Graduated) and N (Did not Graduate).

Summary

This study analyzed 613 undergraduate student records from 2004 to 2018, ensuring a minimum four-year window for graduation assessment, though only 20.6% of students achieved a four-year graduation. The cohort was nearly evenly split between transfer students (48.1%) and first-time freshmen (51.9%). Key demographic information included Pell Grant recipients (47.5%) and first-generation college students (52.9%). The AI Decision Tree Model was used to predict the Four-Year Graduate Rate (Class Y/N). The results identified transfer status as the primary determinant for the initial split. The model immediately established a low overall graduation expectation, with 79.2% of all students predicted not to graduate within four years. The analysis revealed a critical finding: transfer students (52.7% of the sample) showed a substantially lower graduation likelihood, with approximately 96.9% predicted not to graduate. Further splits on Pell status and first-generation status refined these predictions, showing that financially stable, non-first-generation, non-transfer students exhibited the highest likelihood of four-year graduation, while Pell-recipient transfer students represented a subgroup with near-zero predicted four-year completion.

This research addresses a critical gap in the empirical identification and transparent modeling of student characteristics that most strongly predict four-year graduation outcomes in construction management programs through the use of interpretable artificial intelligence methods. The findings demonstrate that decision tree models can clearly identify specific student characteristics associated with increased graduation risk, enabling institutions to proactively implement targeted, data-informed interventions. By aligning institutional support strategies with the most influential variables identified by the decision tree, universities can shift from reactive approaches to proactive, precision-based interventions tailored to distinct student risk profiles. However, the identification and evaluation of specific intervention strategies were beyond the scope of this study; therefore, future research is needed to determine which targeted supports are most effective for addressing the particular characteristics associated with delayed or non-completion of four-year graduation.

A key limitation of this study is its narrow scope, stemming from the dataset's focus on a single university and a specific 14-year time frame (2004–2018), which may limit the generalizability of the findings to other institutions or more recent cohorts. Furthermore, the model is strictly constrained to predicting the stringent four-year graduation rate, thus excluding the large number of students who succeed by graduating within five or six years. Most critically, the AI model relies exclusively on three initial student status flags (transfer, Pell, and first-generation status), omitting other influential variables such as high school GPA, course load, or subsequent institutional academic performance, which could significantly enhance the predictive accuracy and depth of insight.

Future research could focus on three primary avenues to enhance the utility and scope of this predictive AI analysis. First, future work could focus on re-training the model using techniques such as oversampling the 'Y' class or under sampling the 'N' class to force the algorithm to learn the nuanced patterns of the successful student group. Second, the study could be expanded by incorporating a multi-institutional dataset to test the consistency of the current findings regarding transfer and Pell status across different academic environments. Third, the predictive target could be broadened by re-training the AI Decision Tree model to classify five-year and six-year completion rates. This would provide a more holistic view of student success and institutional efficiency. Finally, model sophistication must be increased by integrating richer feature data, such as student High School GPA, university term GPA, and specific course completion metrics, to move beyond initial indicators and capture the dynamic nature of academic performance.

Acknowledgement

The research team would like to thank the support from the University of Texas at San Antonio (UTSA) Institutional Research for providing the data for this research. This publication should not be construed to represent UTSA determination or policy.

References

Alanazi, B. S. (2025). A Comparative Study of Traditional Statistical Methods and Machine Learning Techniques for Improved Predictive Models. *International Journal of Analysis and Applications*, *23*, 18–18. https://doi.org/10.28924/2291-8639-23-2025-18

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: From efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1124553

Burgoon, J., & Elliott, J. (2021). *Evidence of Opportunity Gaps in Construction Management Education*. 559–549. https://doi.org/10.29007/gx1x

Burgoon, J. M., Roebuck Jr, A., & Elliott, J. W. (2024). Evidence of Opportunity Gaps in Construction Education: A Longitudinal Analysis of Student Success. *International Journal of Construction Education and Research*, *20*(1), 83–97. https://doi.org/10.1080/15578771.2023.2186980

Castro-Montoya, B., Vélez-Gómez, P., Segura-Cardona, A., & French, B. F. (2025). A cultural adaptation of Tinto's student integration theory in undergraduate students of a private university in Colombia. *Cogent Education*, *12*(1), 2479384. https://doi.org/10.1080/2331186X.2025.2479384

Collins, C. (2021). *Lateral Transfer Students: What Drives Their Decision to Leave and Why Do They Experience Transfer Shock* [Ed.D., Northeastern University]. https://www.proquest.com/docview/2564492463/abstract/1FAA1A795661466CPQ/1

Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, *56*(5), 4765–4800. https://doi.org/10.1007/s10462-022-10275-5

Cranmer, S. J., & Desmarais, B. A. (2017). What Can We Learn from Predictive Modeling? *Political Analysis*, *25*(2), 145–166. https://doi.org/10.1017/pan.2017.3

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Reports (CS)*. https://scholarworks.utep.edu/cs_techrep/1209

Jimenez-Roa, L. A., Heskes, T., & Stoelinga, M. (2023, October 3). *Fault Trees, Decision Trees, And Binary Decision Diagrams: A Systematic Comparison*. arXiv.Org. https://doi.org/10.3850/978-981-18-2016-8_241-cd

Karimi, H., Taylor, T. R. B., Dadi, G. B., Goodrum, P. M., & Srinivasan, C. (2018). Impact of Skilled Labor Availability on Construction Project Cost Performance. *Journal of Construction Engineering and Management*, *144*(7), 04018057. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001512

Lau, Y., Tang, Y. M., Yiu, N. S. N., Ho, C. S. W., Kwok, W. Y. Y., & Cheung, K. (2022). Perceptions and Challenges of Engineering and Science Transfer Students From Community College to University in a Chinese Educational Context. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.797888

Müller, C. C., Kayyali, M., & Mohamed Elzomor. (2023, June 25). *Factors Affecting On-Time Graduation of Engineering and Construction Management Undergraduates at a Minority Serving Institution*. 2023 ASEE Annual Conference & Exposition. https://peer.asee.org/factors-affecting-on-time-graduation-of-engineering-and-construction-management-undergraduates-at-a-minority-serving-institution

Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *2011 IEEE Control and System Graduate Research Colloquium*, 37–42. https://doi.org/10.1109/ICSGRC.2011.5991826

Olayinka, O. H. (2019). Leveraging predictive analytics and machine learning for strategic business decision-making and competitive advantage. *International Journal of Computer Applications Technology and Research*, *8*(12), 473–486.

Salah Hashim, A., Akeel Awadh, W., & Khalaf Hamoud, A. (2020). Student Performance Prediction Model based on Supervised Machine Learning Algorithms. *IOP Conference Series: Materials Science and Engineering*, *928*(3), 032019. https://doi.org/10.1088/1757-899X/928/3/032019

Sulbaran, T. (2023). Evaluating the Comprehension of Construction Schedules of an Artificial Intelligence. *23° International Conference on Construction Applications of Virtual Reality*.

Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Whitmore, S. R., & Ojajuni, O. P. (2023, June 25). *Are Construction Management Education Programs Producing Sufficient Numbers of Minority Graduates to Meet Demand?* 2023 ASEE Annual Conference & Exposition. https://peer.asee.org/are-construction-management-education-programs-producing-sufficient-numbers-of-minority-graduates-to-meet-demand