Predicting Student Success in Higher Education: A Data-Informed Analysis of Key Institutional Variables

Prabesh Shrestha

Xiaowei Xu

About the Authors

Prabesh Shrestha serves as Executive Director of Institutional Effectiveness and Chief Information Officer at a 2-year college within the Minnesota State system. Xiaowei Xu is professor at the University of Arkansas at Little Rock.

Abstract

Higher education institutions increasingly rely on data-informed strategies to support student success, yet identifying the most impactful variables remains a challenge. This study integrates insights from existing literature with institutional data from a Midwestern, profession-focused associate degree–granting higher education institution to evaluate predictors of academic success, defined as the successful completion of all registered courses in a semester. Using Random Forest and generalized regression models, the analysis reveals that academic standing, early alert indicators (flags), and positive reinforcement (kudos) are the most consistent and significant predictors of student success. Financial aid and advisor contact show moderate, context-dependent effects, while demographic and enrollment characteristics, such as gender, age, and first-generation status, exhibit limited predictive power. These findings underscore the importance of proactive academic support, timely interventions, and recognition systems. Institutions can enhance student success by prioritizing early alerts, targeted advising, and financial support for their students.

Keywords: student success, predictive modeling, Random Forest, generalized regression, early alert indicators, academic interventions, institutional engagement, data-informed decision-making

The AIR Professional File, Spring 2026

Article 182

https://doi.org/10.34315/apf1822026

INTRODUCTION

Higher education institutions are committed to fostering student success, the definition of which varies widely based on individual goals, ranging from simply earning a degree to improving career prospects, expanding networks, or gaining a deeper understanding of the world. To meet these diverse aspirations, institutions must identify and implement effective services and support tailored to their unique student populations. Simultaneously, they must comply with standards set by accrediting bodies and government agencies, which evaluate institutional performance through metrics such as enrollment, course completion, graduation rates, loan default rates, retention, licensure pass rates, employment outcomes, and transfer rates (Denton & Hatfield, 2023).

Given this context, understanding the specific indicators of student success is essential for designing effective interventions. Many institutions have invested in early alert systems and basic needs support to address both academic and nonacademic challenges. However, assessing the impact of these initiatives requires rigorous analysis of their influence on student outcomes. Since institutions collect vast amounts of structured and unstructured data, the challenge lies in identifying which data points most effectively inform strategies for student support, while recognizing that no single variable universally predicts success across all student groups.

This study reviewed existing literature to address inconsistencies in how institutions define student success and identify the variables most associated with it. By examining prevailing definitions of student success, evaluating frequently used variables, and assessing their potential contributions, this study aimed to determine how these constructs can be used effectively to measure and improve student outcomes. The study focused on the following three questions: (1) Which variables most strongly predict semester-level course completion? (2) How do predictor effects vary across semesters? and (3) Do early alerts (flags, kudos, advisor contact) provide predictive value beyond demographics and financial support?

LITERATURE REVIEW

Based on the HLC conference (Denton & Hatfield, 2023), for any institution a strategic approach to student success is to have a mission, understand the student population the mission serves, then align initiatives to support the students, analyze institutional infrastructure, and engage the college community.

Students' demographic data, socioeconomic status, first-generation status, and a few other data points could be helpful to understand and see the changes in student population within the institution. Getting to know students through a survey as they start their higher education can help capture their goals and their reason for choosing a specific institution, as well as to understand what they need to be successful.

An institution can review its initiatives to understand how it is providing support to the students and if those supports and services encompass all its students. Evaluating its policies and practices will help identify the impediments and challenges its students may be facing to succeed. The institution can then identify the area(s) in the institution that is (are) committed to student success, and ensuring that the students agree.

How does an institution start this process? To measure the effectiveness of initiatives on student success, it is important to define student success, identify the metrics used, and measure the effectiveness of the metrics on student success.

This literature review focuses on understanding what student success is and which variables were used to measure student success, then reviews the effectiveness of those variables.

Defining Student Success

Student success has been conceptualized across multiple dimensions, encompassing academic achievement, critical thinking, and social-emotional well-being. Van der Zanden et al. (2018) emphasized these domains, exploring predictors, similarities, differences, and interconnections. Academic achievement is typically measured through GPA, credits earned, and persistence, as noted by Van Rooij et al. (2017), whereas Webb et al. (2020) highlighted the impact of grading scales on fail rates, cautioning against percentage-based systems. In community-college contexts, success has been operationalized as persistence, attainment, and transfer, with Calcagno et al. (2018) and Hensley et al. (2021) linking higher grades and persistence to positive outcomes. McLeod (2019) and Thalluri (2016) identified the critical first semester as pivotal for retention and timely completion, while Gardner and Brooks (2018) emphasized dropout, stop-out, grades, and graduation in the MOOC (massive open online course) context.

Beyond academic metrics, student success is increasingly framed as individualized and socially situated. Lydster et al. (2019) argued that higher education functions as a bridge to professional goals, making success a personal construct. Picton et al. (2018) found that students initially associate success with time management and external feedback; over time, however, behavioral engagement and emotional factors such as belonging and self-efficacy become central. Critically, Weatherton et al. (2021) highlighted systemic inequities in STEM, critiquing narrowly defined institutional measures and advocating for inclusive, nonacademic indicators informed by student perspectives, community-based research, and policy reforms.

Variables Used to Measure Student Success

Variables used to measure students' success span academic, relational, and sociocultural domains. Academic metrics include GPA, course grades, credits, and gateway course completion (Belfield et al., 2018; Chamely-Wiik et al., 2023; Rosser-Majors et al., 2022). Relational and environmental factors such as parental involvement, mentorship, and cultural support programs influence outcomes (Bell et al., 2023; Kranstuber et al., 2012; Lydster & Murray, 2019). Sociocultural adjustment and psychosocial well-being are increasingly recognized as critical for persistence, with interventions including counseling, pre-orientation programs, and first-year experience courses supporting student adaptation (Brunsting et al., 2018; Thalluri, 2016). Meta-analytic evidence suggests that high school GPA, academic motivation, study skills, prior preparation, social support, and program participation strongly predict academic achievement, while inquiry-based learning, challenging pedagogy, and first-year engagement predict critical thinking. Social-emotional well-being is influenced by coping self-efficacy, a sense of belonging, and social relationships, though demographic effects are inconsistent, with first-generation students often at a disadvantage (Van der Zanden et al., 2018). Collectively, these findings underscore the multidimensional nature of student success and the need for integrative, context-sensitive measures.

Table 1. Summary of Variables Derived from the Literature Review, Categorized with References

Variable Category	Variables	Reference(s)
Academic Performance	• Completion of gatekeeper courses in the first year • First-semester GPA • Cumulative GPA • Degree GPA	Belfield et al. (2018) Van der Zanden et al. (2018) Chamely-Wiik et al. (2023)
Time to Completion	• Time to graduate	Chamely-Wiik et al. (2023)
Postgraduation Outcomes	• Postgraduation experience (graduate/professional, undergraduate/certificate/none)	Chamely-Wiik et al. (2023)
Demographics & Background	• Gender • Race/ethnicity • Socioeconomic status • Parental education/income • First-generation status	Calcagno et al. (2008) Van der Zanden et al. (2018) Hensley et al. (2021)
Academic Preparation	• High school GPA • Advanced placement credits • University-level courses in high school	Van der Zanden et al. (2018)
Study Skills & Motivation	• Time spent studying • Study strategies • Academic motivation • Effort	Hensley et al. (2021) Van der Zanden et al. (2018)
Institutional Support & Programs	• Orientation programs • First-year seminars • Tutoring for indigenous students • Study-buddy-support (SBS) scheme	Thalluri (2016) Lydster & Murray (2019) Van der Zanden et al. (2018)
Social & Emotional Well-being	• Psychological well-being • Loneliness • Coping self-efficacy • Sense of belonging	Van der Zanden et al. (2018)
Social Relationships	• Parent-child relationships • Faculty and peer relationships • Family communication	Kranstuber et al. (2012) Van der Zanden et al. (2018)
Institutional Characteristics	• Size • Part-time faculty • Degree mix • Tuition • Expenditure • Location	Calcagno et al. (2008)
Accessibility & Equity	• Accessibility of resources • Study abroad programs for students of color	Betts et al. (2013) Bell et al. (2023)
Feedback & Communication	• Quality and consistency of feedback	Diem & Hairrell (2019)
Course Engagement	• Attendance policies • Selection tests for program entry	Comeford (2022) Rosser-Majors et al. (2022)
Critical Thinking Development	• Pre-university critical thinking • Inquiry-based learning • Challenging questions	Van der Zanden et al. (2018)
Financial Support	• Educational tax benefits	Elsayed (2016)
Parental Influence	• Parents' knowledge • Message characteristics and relational satisfaction	Forster & van de Werfhorst (2019) Kranstuber et al. (2012)

METHODOLOGY

Following the literature review, the next step was to find similar variables from an institution, to identify if those data can be obtained, and then to analyze them to assess their impact on student success.

Since there are multiple definitions of student success, for this research we considered semester-level data, and defined student success as the completion of all the courses students registered in any given semester with a grade of C or better.

The study was completed using data from a public, 2-year, profession-focused institution in the US Midwest (Carnegie: Associate, Small), enrolling around 3,000 degree-seeking students a year; around 23% were students of color and around 45% had a Pell award for fiscal year 2024. We integrated records from the Student Information System (SIS) and the EAB early alert platform (Starfish) for seven semesters, from Spring 2022 through Spring 2024:

- Fiscal year 2022: Spring 2022
- Fiscal year 2023: Summer 2022, Fall 2022, Spring 2023
- Fiscal year 2024: Summer 2023, Fall 2023, Spring 2024

  Records were keyed by unique student identifier and semester. Analyses were conducted at the semester level. The following variables were made available through the institution after institutional research board (IRB) approval:

Table 2. List of Variables with Operational Descriptions and Coding Scheme

Variables	Description	Coding Scheme	
Admission Status	This identifies if the student is a regular student or a transfer student.	Dummy1 (R-T)	Dummy2 (R-PD)
		0: Regular	0: Regular
		1: Transfer	0: Transfer
		0: Previous degree	1: Previous degree
Advisor Contact	This identifies if the student had contact with an advisor after a flag was raised.	1: Had contact with advisor	
		0: Did not have contact with an advisor.	
Age group	This identifies if the students are traditional or nontraditional.	1: Age 25 or older	
		0: Age 24 or younger	
Completion (Student \ success)	This identifies if a student completed all the courses they enrolled in any given semester with C or better.	1: Completed	
		0: Not completed	
First Generation	Neither parent has completed postsecondary education.	1: Yes	
		0: No	
Flag	Students identified to have some concerns in at least one of the classes, with the concern raised by faculty or advisor.	1: Flag raised	
		0: No flag	
Gender	This identifies the gender of the student.	Dummy1 (M-F)	Dummy2 (M-O)
		0: Male	0: Male
		1: Female	0: Female
		0: Other	1: Other
Kudos	Students received kudos (positive feedback) from an advisor or faculty member during the semester.	1: Received kudos	
		0: Did not receive kudos	
Pell Eligibility	This identifies if the student is Pell eligible.	1: Pell eligible	
		0: Unknown	
Student of Color (SOC)	This identifies if the student's race was White or not.	1: Yes	
		0: No	
Referrals	Students were referred to tutors or other student services.	1: Students referred to services	
		0: Students not referred to services	
Scholarship	Received financial assistance that was coded as scholarship by the institution.	1=Received scholarship	
		0=Did not receive scholarship	
Student Status	This identifies if the student is a freshman, sophomore, junior, or senior.	Dummy1 (F-P)	Dummy2 (F-S)
		0: Freshman	0: Freshman
		1: Previous degree	0: Previous degree
		0: Sophomore	1: Sophomore
Term Load	This identifies if the student is taking courses as a full-time or part-time student.	1: Full time	
		0: Part time	

Categorical variables were dummy coded with a single reference category per variable to ensure interpretability and model consistency. Reference categories were Male (Gender), Freshman (Student Status), Regular (Admission Status), No (Pell eligibility, Student of Color, Term Load, Flag, Kudos, Advisor Contact, Scholarship, Referral), Traditional (Age Group). Dummy indicators were created for all nonreference levels. Binary variables were coded 0 = No, 1 = Yes.

Data collected for these variables included all students age 18 and older who were admitted as regular or transfer students, or students with previous degree, between the Spring semester of 2022 and the Spring semester of 2024. The dataset was assembled using fiscal year, semester, and unique student identifiers to define unique records for analysis. Student success was measured using the completion variable.

We conducted range and logic checks across all fields, deduplicated records by student–semester keys, and validated joins across SIS and early alert sources. Missing data rates were reviewed per variable; records with missing values on analysis fields were excluded. Future work will examine missingness patterns (missing completely at random [MCAR]/missing at random [MAR]) and potential exclusion bias.

Since students continue to take courses for multiple semesters, we analyzed each semester's data independently. Given repeated measures across semesters, results are interpreted as exploratory patterns; future validation will require student-level aggregation or mixed-effects with clustered errors. For this study, predictive modeling was preferred to help with the data analysis. Table 3 summarizes the descriptive statistics of the dataset collected as part of this study.

Table 3. Descriptive Statistics of the Study Sample

Terms		Semester 1	Semester 2	Semester 3	Semester 4	Semester 5	Semester 6	Semester 7
Student Count		1,970	387	2,109	1,901	372	2,021	1,829
Term Load	Full Time	59%	53%	61%	60%	54%	61%	60%
Age Group	Non-Trad	33%	33%	32%	34%	36%	32%	34%
Admission Status	Previous Deg	3%	3%	3%	3%	3%	3%	3%
	Transfer	46%	47%	44%	46%	52%	41%	45%
Class	Freshman	26%	21%	36%	22%	17%	36%	24%
	Sophomore	71%	76%	61%	75%	80%	61%	73%
	Previous Deg	3%	3%	3%	3%	3%	3%	3%
Gender	Female	62%	71%	61%	61%	75%	62%	63%
First Gen	Yes	26%	31%	26%	26%	26%	26%	26%
Pell Elig	Yes	47%	42%	46%	46%	47%	47%	47%
Student of Color	Yes	23%	21%	22%	22%	19%	22%	22%
Flag	Yes	24%	12%	26%	27%	16%	7%	25%
Referral*	Yes	4%	1%	4%	4%	0%	0%	5%
Kudos	Yes	54%	36%	43%	44%	29%	1%	48%
Advisor Contact	Yes	42%	26%	50%	58%	25%	29%	2%
Scholarship	Yes	26%	32%	28%	29%	31%	27%	26%
Completion	Yes	77%	92%	77%	78%	92%	76%	80%

Note: * Categories reported as 0% reflect proportions below 1%, despite the presence of corresponding cases in the underlying dataset.

Several intervention-related variables (e.g., referrals, advisor contact, kudos) show unusually low frequencies in select semesters. These patterns are due to staffing changes, system adoption timing, and evolving data-entry practices rather than to the absence of student support activity. Reported values reflect recorded administrative data and may underrepresent informal or undocumented interventions.

This study employed two complementary analytic approaches: Random Forest decision trees and generalized regression (logit function), because each addresses a distinct interpretive objective. Random Forest models were used to capture complex, nonlinear interactions among predictors and to identify combinations of characteristics associated with elevated risk or advantage. By averaging across multiple decision trees, this approach reduces overfitting and provides a robust assessment of relative variable importance. Five-level decision trees were generated to illustrate how predictors differentiate student outcomes and to identify student groups that may benefit from targeted or differentiated interventions.

While Random Forest models are well suited for ranking predictors and identifying conditional risk patterns, they do not provide interpretable probability estimates. To estimate average, population-level effects, generalized regression models with a logit link were applied. Logistic

regression produces odds ratios that quantify the direction and magnitude of association between each predictor and the likelihood of completing all registered courses, holding other variables constant. Using both approaches allows the analysis to integrate segmentation-based insights with population-level inference; differences across methods are therefore expected and substantively meaningful.

RESULTS

The results from Random Forest are shared in Figures 1a through 1g. The performance metrics of the model demonstrated generally strong predictive performance throughout the seven semesters. Its accuracy, which shows how often the model guessed correctly, ranged from 0.764 to 0.947, while F1-scores, a measure that balances correct predictions with avoiding mistakes, ranged from 0.75 to 0.93. Semesters 2 and 5, which had the smallest sample sizes, achieved the highest performance, with both accuracy and F1-scores exceeding 0.93, suggesting that predictions for these cohorts were highly reliable. However, because these groups were small, these very high numbers might be somewhat misleading.

For the larger semesters, including Semesters 1, 3, 4, 6, and 7, accuracy and F1-scores were 0.764–0.831 and 0.75–0.82, respectively, indicating that the model performed consistently even across variables and larger datasets. Semester 6 posed a slightly greater challenge, exhibiting the lowest F1-score and accuracy among the larger cohorts, possibly reflecting increased complexity or variability in the underlying data. Overall, the model demonstrated reliable performance across all semesters, with exceptionally high accuracy in smaller, more-uniform cohorts; the model showed moderately lower, but still robust, predictive power in the larger, more-complex datasets.

In addition, across the seven semesters the five-level decision trees consistently identified a core set of predictors associated with student outcomes. PELL_ELIG_FLAG (Pell Grant eligibility) and scholarship status frequently appeared as root nodes, confirming the strong predictive power of financial aid for academic performance and persistence. Students with financial support were consistently more likely to succeed. CLASS, AGE_GROUP, and GENDER functioned as demographic and academic indicators, generally appearing deeper in the trees and suggesting a secondary, though still relevant, influence. Among these, CLASS (class level) was particularly critical, with freshmen consistently at higher risk compared to students in advanced levels.

Institutional engagement variables—including Flag, Referral, Kudos, and Advisor Contact—also emerged as prominent features, underscoring the influence of early-alert systems and targeted interventions. Flag and Referral reliably identified students experiencing academic difficulty, whereas Kudos were positively associated with persistence and stronger performance. Advisor Contact demonstrated a dual role: in earlier semesters it appeared to facilitate student success, while in later semesters it often reflected students already encountering challenges. The presence of ADM_STAT (admission status) in later models, together with AGE_GROUP and GENDER, further suggests their increasing relevance over time.

Collectively, these findings highlight the dynamic interplay between financial, demographic, and behavioral factors, with financial aid and institutional engagement consistently emerging as the most

significant levers for student success. Importantly, the models also indicated shifting patterns of influence, as factors such as admission status and referrals gained predictive weight in later semesters. Overall, the Random Forest analysis demonstrates that financial support, academic standing, and timely engagement signals are the most powerful drivers of student outcomes, while demographic factors, though present, play a comparatively minor role.

Figure 1a. Random Forest Model Performance for Semester 1 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.7639593908629442

	precision	recall	f1-score	support
0	0.62	0.43	0.51	112
1	0.80	0.90	0.84	282
accuracy			0.76	394
macro avg	0.71	0.66	0.68	394
weighted avg	0.75	0.76	0.75	394

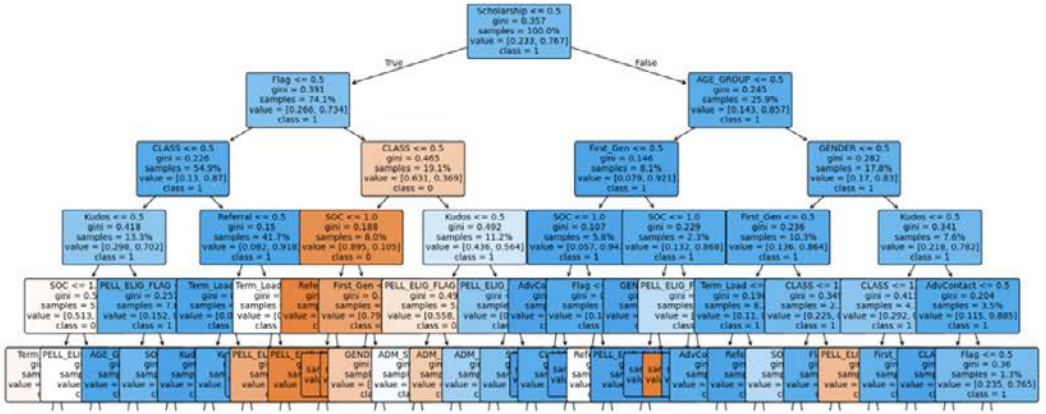


Figure 1b. Random Forest Model Performance for Semester 2 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.9466666666666667

	precision	recall	f1-score	support
0	1.00	0.20	0.33	5
1	0.95	1.00	0.97	70
accuracy			0.95	75
macro avg	0.97	0.60	0.65	75
weighted avg	0.95	0.95	0.93	75

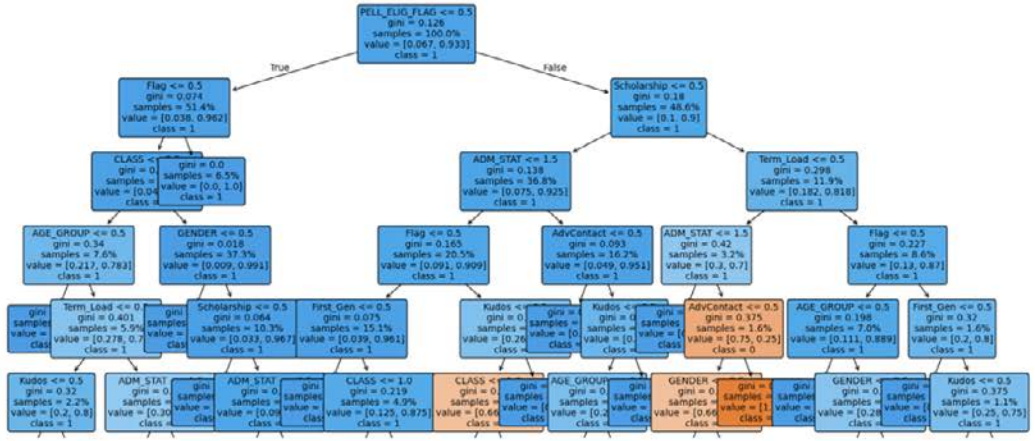


Figure 1c. Random Forest Model Performance for Semester 3 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.8151658767772512

	precision	recall	f1-score	support
0	0.54	0.45	0.49	84
1	0.87	0.91	0.89	338
accuracy			0.82	422
macro avg	0.71	0.68	0.69	422
weighted avg	0.80	0.82	0.81	422

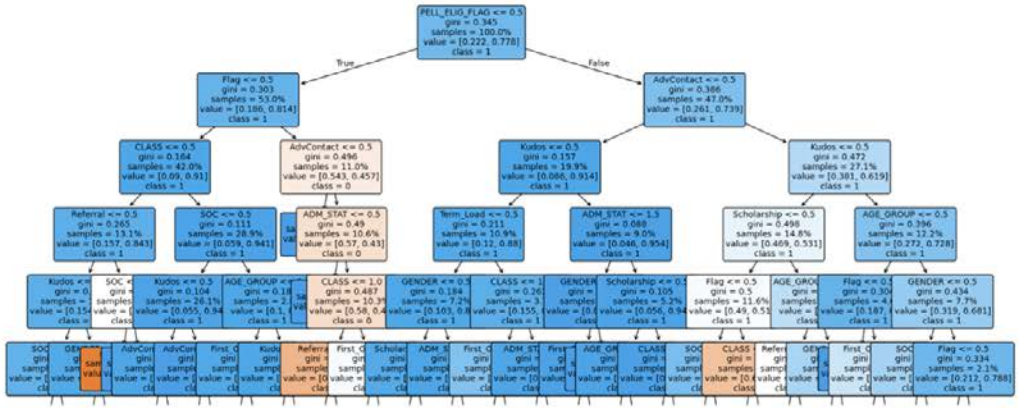


Figure 1d. Random Forest Model Performance for Semester 4 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.8293963254593176

	precision	recall	f1-score	support
0	0.63	0.51	0.56	82
1	0.87	0.92	0.89	299
accuracy			0.83	381
macro avg	0.75	0.71	0.73	381
weighted avg	0.82	0.83	0.82	381

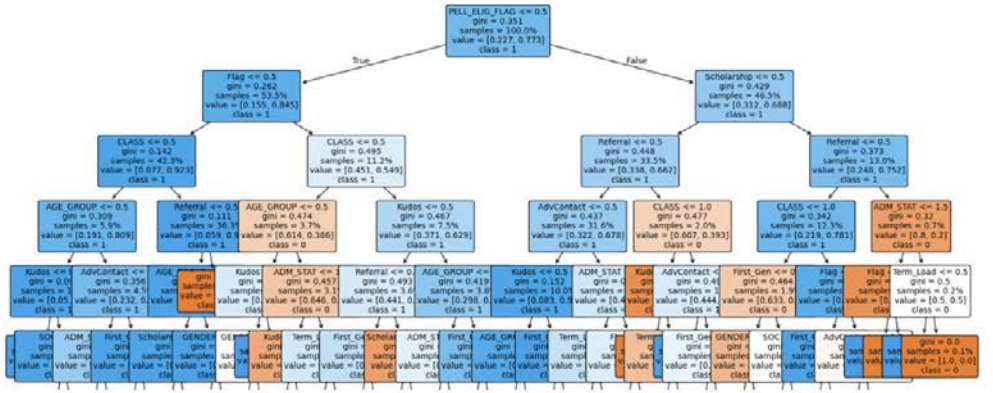


Figure 1e. Random Forest Model Performance for Semester 5 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.9466666666666667

	precision	recall	f1-score	support
0	1.00	0.20	0.33	5
1	0.95	1.00	0.97	70
accuracy			0.95	75
macro avg	0.97	0.60	0.65	75
weighted avg	0.95	0.95	0.93	75

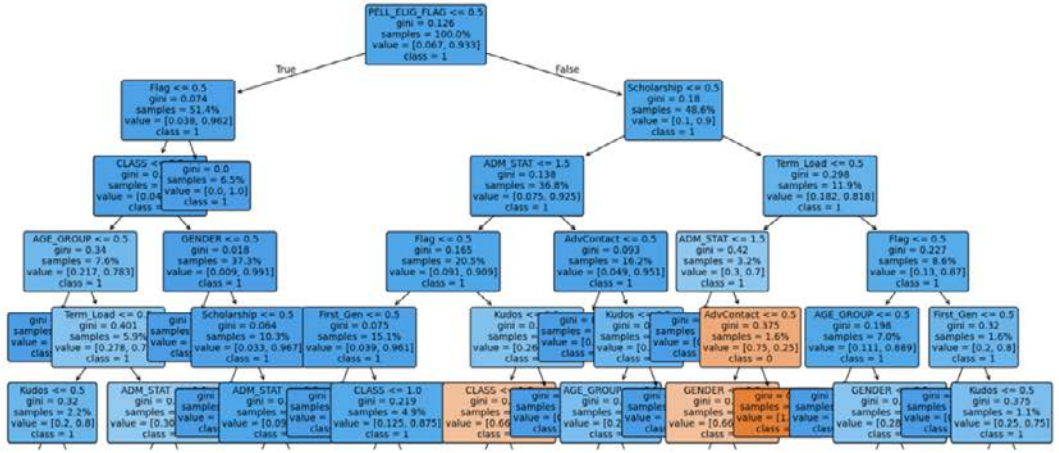


Figure 1f. Random Forest Model Performance for Semester 6 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.7679012345679013

	precision	recall	f1-score	support
0	0.53	0.32	0.40	98
1	0.81	0.91	0.86	307
accuracy			0.77	405
macro avg	0.67	0.61	0.63	405
weighted avg	0.74	0.77	0.75	405

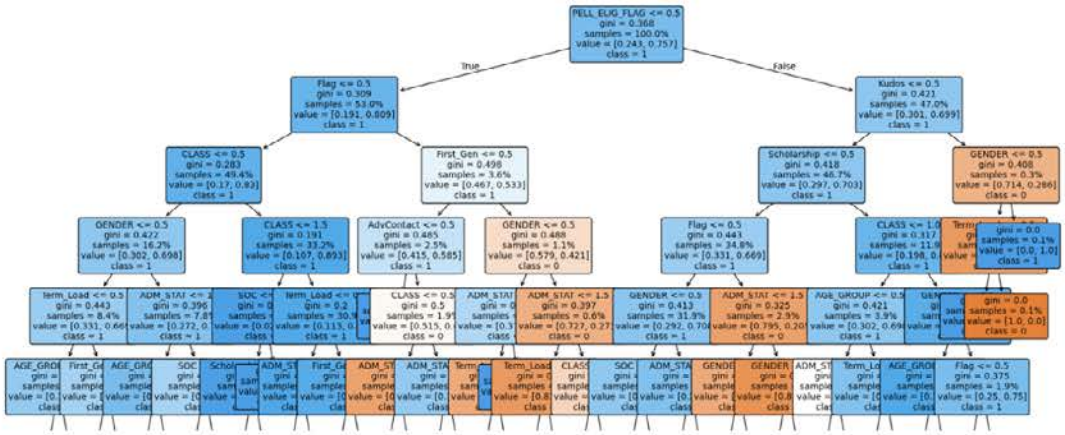
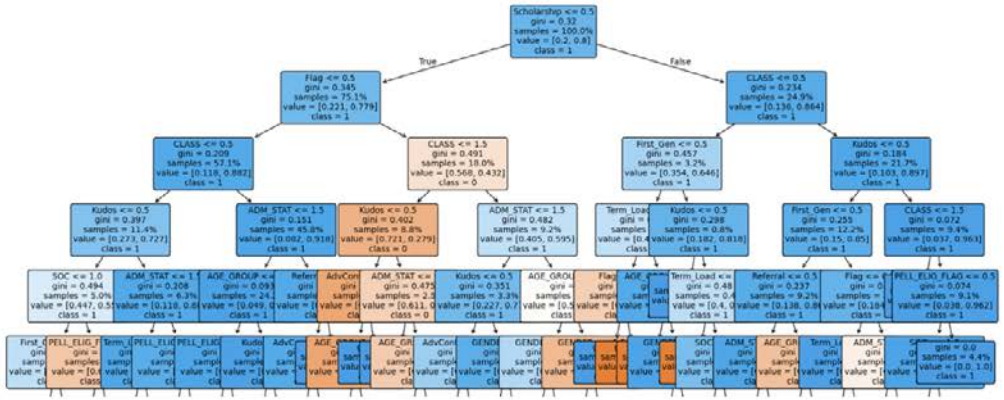


Figure 1g. Random Forest Model Performance for Semester 7 (Precision, Recall, and F1-Score) and Five-Level Decision Tree Depicting Predictor Significance

Accuracy: 0.8306010928961749

	precision	recall	f1-score	support
0	0.57	0.43	0.49	69
1	0.88	0.92	0.90	297
accuracy			0.83	366
macro avg	0.72	0.68	0.70	366
weighted avg	0.82	0.83	0.82	366



Complementing these segmentation-based findings, generalized regression results share significance of predictors and estimate average, population-level associations between predictors and course completion, expressed as odds ratios. The results can be seen in Tables 4 and 5.

Table 4. Statistical Significance (p-value) of Variables by Semester

	p-value						
	Semester 1	Semester 2	Semester 3	Semester 4	Semester 5	Semester 6	Semester 7
Term_Load[F-P]	#	#	#	#	#	#	#
AGE_GROUP[N-T]	#	#	#	#	#	#	#
ADM_STAT[P-T]	#	#	#	#	#	#	#
ADM_STAT[R-T]	#	#	#	#	#	#	#
CLASS[F-S]	***	*	***	***	#	***	***
CLASS[P-S]	.	#	.	#	#	#	#
GENDER[F-O]	#	#	#	#	#	#	#
GENDER[M-O]	#		#	#		#	#
First_Gen[N-Y]	#	#	#	#	#	#	#
PELL_ELIG_FLAG[U-Y]	**	#	#	**	#	***	#
SOC[N-Y]	#	#	#	#	#	*	#
SOC[U-Y]	#					#	#
Flag[N-Y]	***	***	***	***	*	***	***
Referral[N-Y]	#	#	*	*	#	#	*
Kudos[N-Y]	***	#	***	***	**	#	***
Advisor Contact[N-Y]	***	#	***	***	#	***	#
Scholarship[N-Y]	**	#	#	#	#	**	#

Note: # for *p*-value > 0.05, * for *p*-value < 0.05, ** for *p*-value < 0.01, and *** for *p*-value < 0.001.

Conventionally, *p*-value of less than 0.05 is considered significant, and with that definition, data from Table 4 suggests that the most influential predictors of student success are CLASS[F-S] (freshman/sophomore), Flag[N-Y] (No/Yes), Kudos[N-Y] (No/Yes), Advisor Contact[N-Y] (No/Yes), and PELL_ELIG_FLAG[U-Y] (Undefined/Yes) based on their repeated significance over seven semesters. Other factors like Scholarship[N-Y], Referral[N-Y], SOC[N-Y] show significance in some semesters. This suggests their impact may depend on the specific semester or context: impact is not consistent. Finally, factors such as GENDER, First_Gen, AGE_GROUP, Term_Load, ADM_STAT[R-T], CLASS[P-S] do not appear to have much effect on student success in this dataset, since they rarely reached statistical significance.

Table 5. Odds Ratios for 13 Study Variables Across Semesters

	Semester 1	Semester 2	Semester 3	Semester 4	Semester 5	Semester 6	Semester 7
Term_Load							
Full-Time/Part-Time	1	0.6234	1	0.8559	1	1	1.3122
AGE_GROUP							
Non-Trad/Trad	1	0.6581	0.8519	0.8910	1.1825	1	0.9339
ADM_STAT							
Previous Degree/Regular	0.4051	1	1.3946	1	1.6160	1	1
Transfer/Regular	1	1	1	1	1.6160	1	1
CLASS							
Previous Degree/Freshman	3.5839	2.5846	1.8229	3.2090	1.7002	2.7836	3.1194
Sophomore/Freshman	3.6664	2.5846	1.8137	3.2090	1.7002	2.7836	3.1194
GENDER							
Female/Male	1.1313	1.8584	1	1	1	1.0474	1
First_Gen							
No/Yes	1.0434	1	1.2211	1	1	1.0296	1
Pell_Elig_Flag							
Pell Elig/Undefined	0.6493	1	0.9397	0.6768	1	0.6877	0.8762
SOC							
Yes/No	0.8513	1	1	0.9835	0.8277	0.7619	0.7625
Flag							
Yes/No	0.1759	0.0922	0.2018	0.1783	0.3048	0.5013	0.1093
Referral							
Yes/No	0.8515	0.2433	0.6270	0.5379	0.2784	1	0.5529
Kudos							
Yes/No	2.5389	1.1239	2.1076	1.6541	3.1974	1	2.4117
Advisor Contact							
Yes/No	0.4499	1	0.3449	0.3661	0.4933	0.4895	1.1210
Scholarship							
Yes/No	1.6402	1	1.1622	1.2874	1	1.4249	1.1413

Key findings are as follows:

- Sophomores have 3.12× the odds of completing all courses relative to freshmen (Semester 7; similar magnitude across terms).
- A flag is associated with ~0.11–0.50× the odds of completion (strong negative predictor across all terms).
- Kudos are linked to ~1.65–3.20× higher odds of completion.

The analysis of odds ratios across all seven semesters provides a detailed view of the factors influencing student success. Across semesters, certain variables consistently emerge as strong predictors, while others show moderate or inconsistent effects.

Class level is a particularly strong and consistent predictor. Freshmen consistently exhibit lower odds of success compared to sophomores or students with previous degrees, indicating that prior academic experience positively influences student outcomes. Similarly, the presence of a flag is a very strong negative predictor across all semesters, suggesting that students identified as needing intervention are at substantially higher risk of lower academic performance. Positive reinforcement measures, such as kudos, consistently correspond to higher odds of success, highlighting the importance of recognition in motivating and supporting students.

Other variables show more moderate or context-dependent effects. Referrals, which indicate that a student was directed toward additional support services, generally correspond to lower odds of success, though the magnitude of this effect varies across semesters. Scholarship receipt demonstrates a modest positive impact, particularly in the early to mid-semesters, suggesting that financial support facilitates better academic outcomes. Likewise, advisor contact is mostly associated with improved odds of success, although a slight reversal is observed in Semester 7, potentially reflecting targeted intervention for students who are already struggling.

In contrast, several demographic and enrollment-related variables, including age group, gender, first-generation status, term load, Pell eligibility, and admission status, exhibit weak or inconsistent effects. Their odds ratios are frequently close to 1, indicating minimal direct influence on student success across semesters.

DISCUSSIONS

Viewed together, the Random Forest and generalized regression results provide a complementary understanding of student success that integrates conditional risk patterns with population-level associations. The Random Forest analyses highlight how combinations of financial context, academic standing, and engagement signals differentiate groups of students with varying likelihoods of course completion, offering insight into which students may benefit from targeted or differentiated interventions. The regression results clarify which factors are most consistently associated with student success across the population, underscoring the strong positive association of academic standing and positive reinforcement (kudos), as well as the substantial negative association of academic flags.

Support-related variables such as referrals, scholarships, and advisor contact demonstrate context-dependent associations, with effects varying across semesters. These patterns suggest that the

timing and circumstances of interventions may influence their relationship with student outcomes. Consistent with prior research, demographic and enrollment characteristics exhibit comparatively weak and inconsistent associations with semester-level course completion in this institutional context.

Overall, differences between the two analytic approaches reflect their distinct interpretive lenses rather than a contradiction. Decision trees emphasize segmentation and conditional risk, while regression models summarize average effects across the student population. Together, the findings reinforce the importance of early identification of academic difficulty, recognition of positive academic behaviors, and structured support (particularly for freshmen) as central components of effective student success strategies.

CONCLUSION

This study underscores the importance of data-informed strategies in predicting and enhancing student success in higher education. Through a comprehensive literature review and analyses using Random Forest and generalized regression models, it identifies academic standing, early alert indicators (e.g., flags), and positive reinforcement (e.g., kudos) as the most consistent and impactful predictors of student success. Financial support mechanisms, including Pell eligibility and scholarships, along with institutional engagement through advisor contact and referrals, also play meaningful—though context-dependent—roles in influencing outcomes.

Conversely, demographic and enrollment characteristics such as gender, age group, first-generation status, and term load showed limited predictive power, suggesting that, while these factors can shape student experiences, they are less effective in forecasting academic performance within the studied context.

These findings point to actionable opportunities for institutions to strengthen student success initiatives. Expanding access to financial aid and scholarships can directly improve persistence, particularly among first-year students. Targeted support for freshmen is especially impactful, given their consistently higher risk of academic difficulty. The demonstrated effectiveness of early-alert mechanisms underscores the need for robust systems to monitor and respond to student progress, while advisor contact should be strategically managed to serve as proactive guidance rather than merely reactive support. Additionally, the growing influence of admission status and referrals in later semesters highlights the importance of ongoing monitoring beyond the first year.

Overall, by prioritizing proactive academic support, timely interventions, and recognition systems, higher education institutions can align resources with evidence-based predictors of success. Such alignment enables institutions to focus on the most influential drivers of student achievement while continuing to address broader demographic and academic factors, ultimately improving persistence and overall student outcomes.

LIMITATIONS OF THE STUDY

This study encountered some limitations that may affect the generalizability of its findings. Key variables, including Satisfactory Academic Progress (SAP) status, instructor-specific information, and detailed student service interactions, were excluded due to data access restrictions and privacy regulations. Moreover, the analysis was confined to a

3-fiscal-year period, which might not capture longer-term trends or variations. Some variables also lacked sufficient data for robust statistical inclusion.

The study relies on observational institutional data, introducing potential selection bias. For instance, students who received advisor contact or early alerts may differ systematically from those who did not, in ways not captured by the available variables (e.g., motivation, prior academic preparation). These unmeasured confounders can influence both the likelihood of intervention and academic outcomes. While predictive models such as Random Forest are robust to certain assumption violations, logistic regression requires independence of observations and correct specification of relationships among predictors—assumptions that were not formally tested here. Sensitivity analyses were not conducted because the primary aim of this study was exploratory: to identify predictive patterns rather than to establish causal relationships. Future research should include diagnostic checks for multicollinearity, goodness-of-fit, and independence, and should also consider advanced approaches such as mixed-effects models or propensity score adjustments to mitigate bias and validate findings.

Given these limitations, the results should be interpreted as preliminary. Future research should incorporate a broader range of variables, extend the analytical timeframe, and include data from multiple institutions to strengthen and validate findings. Investigating potential confounding factors and expanding institutional contexts will enhance the reliability and applicability of predictive models in higher education.

REFERENCES

Belfield, C. R., Crosta, P., & Jenkins, D. (2018). Gatekeeper course completion and student success. *Community College Review, 46*(1), 23–42.

Bell, A., Smith, J., & Johnson, L. (2023). Study abroad programs and student success among students of color. *Journal of Higher Education Diversity, 16*(2), 45–62.

Betts, K., Cohen, A. H., Veit, D. P., Alphin, H. C., Broadus, C., & Allen, D. (2013). Strategies to Increase Online Student Success for Students with Disabilities. *Journal of Asynchronous Learning Networks, 17*(3), 49–64. https://eric.ed.gov/?id=EJ1018265

Brunsting, N., Sreckovic, M., & Lane, K. (2018). Supporting socio-cultural adjustment in first-year students: Counseling, mentoring, and orientation programs. *Journal of College Student Development, 59*(3), 289–306.

Calcagno, J. C., Crosta, P. M., Bailey, T. R., & Jenkins, D. (2008). The impact of institutional and individual variables on community college outcomes. *Educational Evaluation and Policy Analysis, 30*(2), 75–95.

Calcagno, J. C., Crosta, P. M., Bailey, T. R., & Jenkins, D. (2018). Persistence and attainment across community colleges: Patterns and predictors. *Research in Higher Education, 59*(4), 456–479.

Chamely-Wiik, D., Thomas, P., & Rojas, L. (2023). Postgraduate experience and success outcomes in higher education. *Higher Education Research & Development, 42*(1), 102–118.

Comeford, M. (2022). Clear course policies and academic success: Implications for student engagement. *College Teaching, 70*(1), 12–22.

Denton, J., & Hatfield, S. (2023, March 25). *Supporting Student Success* [Workshop presentation]. HLC Conference, Chicago.

Diem, A., & Hairrell, A. (2019). Feedback clarity and cognitive development in college students. *Assessment & Evaluation in Higher Education, 44*(6), 867–882.

Elsayed, W. (2016). Educational tax benefits and college completion rates. *Economics of Education Review, 50*, 1–12.

Forster, A., & van de Werfhorst, H. (2019). Parental knowledge as a predictor of student success. *Sociology of Education, 92*(3), 235–255.

Gardner, J., & Brooks, C. (2018). Measuring success in MOOCs: Dropout, stop-out, grades, and graduation. *Online Learning Journal, 22*(2), 1–21.

Hensley, L., Kim, S., & Martinez, R. (2021). Study strategies and persistence in biology courses. *Journal of STEM Education, 22*(4), 34–47.

Kranstuber, H., Carr, P., & Hosek, A. (2012). Parental influences and college completion outcomes. *Journal of College Student Retention, 14*(3), 257–276.

Lydster, J., & Murray, R. (2019). Indigenous tutoring centers and student success. *Journal of College Student Development, 60*(5), 627–642.

Lydster, J., Smith, K., & Johnson, M. (2019). Individualized perspectives on higher education success. *Higher Education Research & Development, 38*(7), 1451–1466.

McLeod, R. (2019). Semester-to-semester retention and institutional strategies. *Community College Journal of Research and Practice, 43*(9), 678–692.

Picton, C., Matthews, J., & McLaughlin, E. (2018). Student perceptions of success: Behavioral and emotional engagement. *Studies in Higher Education, 43*(9), 1685–1700.

Rosser-Majors, B., Nguyen, H., & Patel, S. (2022). Gateway courses and predictive indicators of academic success. *Journal of College Academic Support, 4*(2), 45–63.

Thalluri, J. (2016). First-semester retention and study buddy support programs. *Journal of College Student Retention, 18*(4), 395–412.

Van der Zanden, P., Denessen, E., Cillessen, A., & Meijer, P. (2018). Student success across academic, critical thinking, and social-emotional domains: A meta-analysis. *Studies in Higher Education, 43*(10), 1701–1717.

Van Rooij, E., Luszczynska, A., & Kommers, P. (2017). Academic persistence and first-semester success. *Educational Research Review, 20*, 47–60.

Weatherton, M., Dawson, S., & Thompson, R. (2021). STEM student success through CRT and inclusive frameworks. *Journal of Diversity in Higher Education, 14*(2), 121–138.

Webb, T., Hunter, A., & Foster, B. (2020). Grading scales and student outcomes: Effects on fail rates. *Assessment in Education: Principles, Policy & Practice, 27*(3), 307–324.