The Impact of Financial Aid and Academic Pathways on Student Success: Analytics, Predictive Modeling, and Interactive Tools

Chaturya Yarradoddi
MPS in Data Science
UMBC
Email: my81293@umbc.edu

Anupam Joshi
CSEE Department
UMBC
Email: joshi@umbc.edu

***Abstract*—The research creates an integrated predictive modeling system which studies how financial assistance together with academic readiness and enrollment methods determine student success in completing their undergraduate studies. The research combines data from 25,926 students through their high school records, placement tests, standardized test results, Advanced Placement coursework, demographic details, and multiple years of financial aid records. The pre-enrollment and first-term variables underwent multiple stages of data preparation which included cleaning, leakage prevention, feature engineering, and deduplication before machine learning model training. The AUC reached 0.846 on the test set after the XGBoost and gradient boosting models received their optimal performance settings. The analysis of model interpretability revealed that financial support amount, mathematics and English placement results, high school academic performance, and transfer student status determined graduation success. The final model is implemented into a Power BI dashboard through Python-based inference which enables users to run real-time predictions, adjust classification thresholds, and test different scenarios. The research developed an operational evidence-based system that institutions can use to make decisions about financial aid distribution, academic readiness programs, and student achievement strategies.**

I. Introduction

Universities now focus on improving graduation rates because they face increasing financial challenges and must meet rising accountability standards. The ability to predict student graduation success enables universities to direct their resources effectively while improving advising services and creating better intervention programs. The system needs to combine different data sets which include academic background information and financial assistance and student characteristics and enrollment methods but must protect against data exposure and maintain uniform data processing.

Financial aid serves as a proven element which determines student ability to stay in school and finish their degree. Research indicates that financial grants and subsidies enable better college entry and graduation success for students from low-income backgrounds and underrepresented groups [1], [2]. Research findings show that students who receive more financial aid tend to achieve better graduation and retention outcomes [3]. Academic readiness stands as a vital element which matches the importance of financial aid. Students who demonstrate better academic readiness through their high school grades and placement test results and Advanced Placement courses tend to achieve higher degree completion rates regardless of their demographic background [4], [5].

The current research provides essential descriptive findings about graduation patterns but institutions require predictive systems which generate personalized probability estimates for operational use. The research solution creates a single analytical database to train machine learning models on student information before enrollment and their first academic term data. The research identifies graduation predictors through model development before implementing the finished model into an interactive Power BI dashboard. The system generates immediate predictions and threshold-based results and simulation tools which help organizations make data-driven financial aid choices and academic support plans.

II. Related Work

Research findings about student achievement demonstrate that financial backing along with academic success and behavioral conduct patterns determine students' ability to complete their degrees. Students who receive adequate need-based or merit-based financial assistance tend to complete their college education. Dynarski [1] examined national financial aid systems to prove that grants and subsidies effectively boost enrollment rates and graduation success for students from low-income backgrounds. The Pell Institute studied institutions that serve students from low-income backgrounds and discovered that their graduation rates vary widely because the way they distribute financial aid affects student retention [2]. The most recent evidence [3] demonstrates that students who receive more financial support tend to stay in college longer and achieve better graduation results while facing lower chances of leaving school.

Academic preparation serves as a primary factor which determines student success in college. Students who achieve high grades in high school and perform well on placement tests and complete challenging courses demonstrate their readiness for college. Research on Advanced Placement (AP) shows that students who take and pass AP exams achieve better graduation results. The research by Dougherty et al. [4] established that students who take AP classes will earn their bachelor's degree even when researchers account for student background characteristics and school environments. The research conducted by Challenge Success [5] demonstrated that students who take AP classes show better academic performance during their first year at college.

Research indicates that students who transfer to four-year institutions achieve academic success at levels comparable to or surpassing those of first-time freshmen. The Government Accountability Office (GAO) [6] determined that students who transfer from community colleges to four-year universities achieve better graduation rates than first-time students at these institutions. The National Center for Education Statistics (NCES) [7] shows that students who bring substantial college credits to four-year institutions achieve better six-year graduation rates than first-time students at these institutions. The Public Policy Institute of California (PPIC) [8] also reported that transfer students who complete lower-division preparation graduate at equal or higher rates than native freshmen, especially within large public university systems. The Community College Research Center (CCRC) [9] published open-access research which confirms that students who transfer with proper preparation and follow specific pathways achieve graduation rates that match or exceed those of traditional first-year students. The research findings from this study match previous studies because transfer student characteristics in the model consistently generated higher graduation probability estimates.

Higher education institutions now use predictive analytics to identify students who need academic or financial support. Most predictive models use restricted data sources while depending on student behavior that has not yet occurred and they do not provide clear policy-relevant insights. The research builds upon previous work by uniting high school records with AP coursework and placement exam results and demographic information and multiple years of financial aid data into a single modeling system. The research enhances both methodological precision and operational decision-making for student success through its implementation of strict data protection measures and interpretability assessments and Power BI deployment.

III. Data and Preparation

The research used financial data from two main sources which included Financial Aid Student Success dataset and Advanced Placement (AP) coursework dataset. The two files underwent initial evaluation before being combined to build an analytical database which contained student information. The Financial Aid file contained 25,960 records which included student demographics and high school data and test results and financial aid information for six years. The AP file contained 58,246 exam records which were combined to show each student's complete AP testing history.

The financial aid dataset allowed the creation of multiple variables which showed the amount and pattern of financial assistance students received from their first day to graduation. The *TotalSupport* variable combined all yearly grant and merit amounts into a single value. The *Supported* variable showed whether students received any financial aid during their studies. The *SupportBin* variable divided TotalSupport into specific ranges which started at $5K and ended at $>20K. . The *NeedStatus* variable identified students who qualified for need-based FAFSA assistance through their legitimate need-based FAFSA entries during any year. The system classified students as non-need when their FAFSA status showed "No FAFSA" or "Unknown" or remained blank. The combination of these features enables complete financial support pattern assessment.

The AP exam data underwent processing to create academic preparation metrics. The AP test descriptions enabled the assignment of AP subjects to seven academic domains which included math and STEM and English and social sciences and language and computer science and arts. The system generated multiple student AP exam metrics for each student including their total exam count and their unique subject count and their earned credits and transferred credits and their highest and average AP scores and their exam distribution across different subjects. The two STEM-focused preparation metrics combined the total number of STEM AP tests with the percentage of all AP tests that belonged to STEM subjects. The variables establish a wide range of student pre-college academic challenge assessment.

The student identifier fields received standardized treatment before dataset combination to prevent numeric-string data type conflicts. The system executed a left join operation based on student ID which produced 25,926 complete student records. Duplicate entries resulted from repeated AP test results and small administrative variations during the merging process. These issues were solved by selecting the highest numeric values and uniting distinct categorical data points for each student record, ensuring one complete record per student.

The system replaced missing numeric data with zeros because these fields represented student counts and credits and financial values. Categorical fields were marked as "Unknown." After imputation, the dataset contained no missing values. The modeling dataset received all its features from the merged dataset except for student progression indicators and multi-year aid patterns and term-by-term enrollment data. The modeling dataset consisted of 33 pre-enrollment and first-term attributes derived from the full 157-feature dataset.

The integrated dataset contains all necessary information about financial support and high school preparation and AP coursework and placement tests and standardized tests and demographic data and enrollment details after completing all data cleaning and duplicate removal and leakage protection steps. The dataset provides the essential data for executing predictive models and policy evaluations in the following sections.

IV. Modeling Framework

The research used student information from before enrollment and their first term to develop predictive models which estimated student graduation chances. The target variable received a value of 1 when students graduated and 0 when they did not complete their studies. The modeling dataset consisted of 33 variables which combined high-school preparation data with placement exam scores and AP indicators and demographic information and residency details and matriculation status and first-term enrollment information and summary financial aid characteristics.

The data received a 70/15/15 stratified distribution for training and validation and testing purposes to maintain the original graduate and non-graduate ratios in all data segments. All models were implemented through scikit-learn pipelines which maintained consistent preprocessing methods and blocked any potential information disclosure. Numerical imputation used median values while z-score normalization was applied for standardization. Categorical variables received most-frequent imputation and one-hot encoding with `handle_unknown=``ignore''`. The `ColumnTransformer` executed these transformations to ensure alignment throughout preprocessing and model training.

The evaluation process included seven machine-learning algorithms which were logistic regression and random forest and gradient boosting and XGBoost and linear SVM and RBF SVM and Gaussian naïve Bayes. The evaluation of model performance used accuracy and recall and F1-score and area under the ROC curve (AUC) metrics on both validation and test datasets. The AUC metric served as the primary evaluation criterion because it operates independently of classification thresholds and performs well under unbalanced data distributions.

The research team performed two sets of experiments: baseline model development followed by PCA-based dimensionality reduction. Numerical variables underwent scaling before PCA transformation. PCA generated new feature spaces but ultimately failed to improve performance for any model. The AUC values decreased after PCA transformation, especially for tree-based models, because PCA removed essential nonlinear patterns required by ensemble algorithms. PCA also reduced interpretability, leading to its removal from the final modeling workflow.

Tree-based models received improved performance through `RandomizedSearchCV` with cross-validation which optimized key hyperparameters including number of estimators, learning rate, maximum depth, subsampling rate, and minimum samples per leaf. These optimization procedures demonstrated that tree-based models excel at identifying complex interactions among academic readiness and financial support variables and demographic factors.

The optimized tree-based models achieved the highest results among all models. The optimized XGBoost model achieved the highest AUC score of 0.846 on validation and test data while maintaining equal precision and recall values. Gradient boosting models showed strong but slightly lower results, logistic regression produced stable moderate performance, and naïve Bayes failed due to its independence assumptions. Ensemble tree models therefore produced the best results for student-success prediction.

The final model selection for feature analysis and partial dependence exploration and scenario simulation and Power BI integration used the optimized XGBoost model because of its superior performance, stability, and interpretability compatibility.

V. Results

A. Model Performance

TABLE I
PERFORMANCE OF MACHINE LEARNING MODELS

Model	Accuracy	Precision	Recall	F1 Score	AUC
XGBoost (Tuned)	0.7578	0.7400	0.7869	0.7627	0.8456
XGBoost (Base)	0.7555	0.7412	0.7770	0.7587	0.8454
Gradient Boosting (Tuned)	0.7511	0.7388	0.7687	0.7534	0.8449
Gradient Boosting (Base)	0.7501	0.7392	0.7646	0.7517	0.8411
Random Forest (Tuned)	0.7503	0.7288	0.7890	0.7577	0.8388
XGBoost (PCA)	0.7490	0.7354	0.7698	0.7522	0.8348
Gradient Boosting (PCA)	0.7372	0.7294	0.7453	0.7373	0.8259
Random Forest (Base)	0.7418	0.7268	0.7661	0.7460	0.8251
Random Forest (PCA)	0.7377	0.7209	0.7666	0.7431	0.8149
Logistic Regression (Tuned)	0.7303	0.7277	0.7266	0.7272	0.8099
Logistic Regression (Base)	0.7308	0.7292	0.7251	0.7271	0.8096
Logistic Regression (PCA)	0.7298	0.7270	0.7266	0.7268	0.8075
GaussianNB (PCA)	0.4983	0.4964	0.9693	0.6566	0.5280
GaussianNB (Base)	0.5014	0.4980	0.9756	0.6594	0.5118

The evaluation of model performance occurred through accuracy, precision, recall, F1 score, and area under the ROC curve (AUC) assessments on the reserved test data. The XGBoost model with optimized parameters produced the highest performance results according to Table I because it reached 0.8456 AUC and 0.7578 accuracy and achieved the highest F1 score among all models. The tuned XGBoost classifier demonstrates the highest discrimination power between graduates and non-graduates while maintaining equal precision and recall performance.

The base XGBoost model and tuned gradient boosting model showed equivalent performance because they both reached AUC values exceeding 0.844. The random forest models showed competitive results but their performance remained slightly lower than the boosting techniques. The AUC values from logistic regression models stayed around 0.81 for all variants while showing stable yet inferior performance. PCA-based preprocessing resulted in decreased predictive accuracy because it eliminated essential information from the original feature domain.

The Gaussian naïve Bayes model produced the worst results because its AUC values reached 0.51–0.53 and its recall values were highly unbalanced due to its inability to model nonlinear relationships present in the data. Overall, ensemble boosting approaches demonstrated superior performance for graduation outcome prediction, with the tuned XGBoost model selected for further analysis and Power BI implementation.

B. Feature Importance

TABLE II
FEATURE IMPORTANCE COMPARISON

Feature	XGB Base	XGB Tuned	GB Tuned
ALEKSScore	0.051	0.047	0.218
AP_ct_computer	0.022	0.020	0.017
AP_ct_math	–	–	0.006
AP_ct_social	–	–	0.005
AP_max_score	0.019	0.019	0.022
AP_total_transfer_credits	–	0.016	0.008
AlgSCORE	0.016	0.016	0.011
CalScore	–	–	0.010
EngSCORE	0.115	0.149	0.171
HS PercentileDesc Unknown	0.019	0.015	–
HighSchool GPABand 2.5–2.99	0.016	–	–
HighSchool GPABand Unknown	0.017	0.017	–
HighSchoolGpa	–	–	0.040
HighSchool RankPercentile	–	–	0.023
MatricGender IPEDS Male	–	–	0.005
Ethnicity Black/African Am.	0.018	0.014	0.007
Ethnicity Hispanic/Latino	0.016	–	–
Residency Out of State	0.016	–	0.006
MatricStatus New Transfer	0.053	0.062	0.043
SATMathScore	0.018	0.019	0.033
SATReadingWritingScore	0.021	0.021	0.051
Sem1 FTPT PT	0.025	0.021	0.013
SupportBin $15K–$20K	–	0.019	–
SupportBin 5K–10K	0.029	0.027	–
SupportBin < 5K	0.042	0.034	0.014
SupportBin > 20K	0.061	0.099	–
Supported Yes	0.018	0.017	–
TotalSupport	0.030	0.026	0.235

The three models demonstrate that graduation results stem from identical fundamental elements. The *EngSCORE* variable produced the strongest connection to student success, followed by *ALEKSScore*, *SATReadingWritingScore*, and *SATMathScore*. Students who received more than $20,000 in financial support (*SupportBin >20K* and *TotalSupport*) tended to stay in school at substantially higher rates. Transfer status (*New Transfer*) appeared as a major structural factor across all three models. Academic readiness, financial stability, and entry pathways served as the essential elements that determined graduation performance.

C. Categorical Gap Analysis

TABLE III
CATEGORICAL GRADUATION PROBABILITY COMPARISONS

Feature	Category	XGB Base	XGB Tuned	GB Tuned
MatricStatus	New Transfer	0.548	0.543	0.548
MatricStatus	New Freshman	0.458	0.460	0.456
Gender	Female	0.523	0.519	0.520
Gender	Male	0.476	0.477	0.476
Ethnicity	American Indian/Alaska Native	0.660	0.641	0.612
Ethnicity	Not Specified	0.600	0.589	0.606
Ethnicity	White	0.536	0.536	0.537
Ethnicity	Asian	0.510	0.505	0.510
Ethnicity	Black/African American	0.428	0.432	0.425
Ethnicity	Hispanic/Latino	0.417	0.428	0.416
Sem1 FT/PT	Full-Time (FT)	0.499	0.498	0.498
Sem1 FT/PT	Part-Time (PT)	0.424	0.428	0.423
Residence	On Campus	0.505	0.503	0.501
Residence	Commuter	0.485	0.486	0.486
HS GPA Band	No GPA	0.641	0.559	0.623
HS GPA Band	3.5–3.99	0.477	0.474	0.475
HS GPA Band	4.0–4.49	0.467	0.466	0.464
HS GPA Band	2.5–2.99	0.462	0.463	0.461
HS Percentile	91–100%	0.633	0.629	0.621
HS Percentile	81–90%	0.536	0.526	0.528
HS Percentile	Unknown	0.482	0.482	0.482
Need Status	Yes	0.503	0.504	0.503
Need Status	No	0.449	0.444	0.450
SupportBin	> 20K	0.682	0.681	0.679
SupportBin	15K–20K	0.587	0.575	0.586
SupportBin	No Support	0.496	0.493	0.495
SupportBin	10K–15K	0.406	0.407	0.403
SupportBin	5K–10K	0.312	0.320	0.314
SupportBin	< 5K	0.258	0.266	0.264

The XGBoost Base, XGBoost Tuned, and Gradient Boosting Tuned models produce identical categorical patterns, demonstrating reliable and consistent results across all three algorithms. The models predict that transfer students achieve the highest graduation probabilities, exceeding the rates of first-time freshmen. The results indicate that students who enter college with completed college credits and prior academic experience show stronger academic momentum and higher completion likelihood. Female students consistently achieve slightly higher predicted outcomes than male students, although the difference remains small.

The models produce identical racial and ethnic distribution patterns. American Indian/Alaska Native students and those who did not specify their ethnicity receive the highest predicted graduation probabilities, while Black/African American and Hispanic/Latino students receive lower predictions. The consistent ordering across all three models suggests that demographic disparities reflect persistent structural patterns present in the underlying dataset rather than variations in algorithmic behavior.

Distinct patterns appear across enrollment characteristics as well. Full-time enrollment yields higher predicted graduation probabilities than part-time enrollment, reflecting the academic stability associated with full course loads. Students who live on campus show slightly better outcomes than commuters, although this difference is minimal. High school preparation follows a clear hierarchy: students in the highest percentile band (91–100%) consistently receive the strongest predictions, followed by middle-percentile groups, while lower-performing or missing-percentile groups receive weaker predictions.

Financial support produces the largest categorical gap across all three models. Students receiving more than $20K in institutional aid achieve predicted graduation rates around 0.68, while students receiving less than $5K fall between 0.26 and 0.32. The stability of these gaps across all three models reinforces the central role of financial aid in student retention and graduation success.

The models also predict unusually high graduation probabilities for students without a reported high school GPA and for students who did not receive institutional support. Further analysis explains these patterns. Most students missing GPA information are transfer students whose previous college performance provides stronger indicators of academic readiness than high school data. Likewise, students who receive no support tend to fall into low-need or financially independent categories and often show strong academic profiles, including high placement scores in English, mathematics, and algorithm-based assessments. Many of these students are concentrated in structured and high-performance majors such as Computer Science, Engineering, Biological Sciences, and Information Systems. Their academic strength and program structure mitigate the absence of financial support and high school GPA data.

Across all categorical dimensions, the three models demonstrate perfect agreement in the ordering and magnitude of effects. This alignment indicates that the observed patterns are fundamental characteristics of the student population and not artifacts of any specific modeling approach.

D. Policy Simulation

TABLE IV
POSITIVE POLICY EFFECTS ON PREDICTED GRADUATION PROBABILITY

Policy Change	XGB Base	XGB Tuned	GB Tuned
Force = New Transfer	+7.62	+5.90	+8.97
Ethnicity = International	+4.84	+2.97	+4.62
Ethnicity = Not Specified	+4.16	+2.74	+4.46
NeedStatus = No	+4.14	+2.86	+0.42
HighSchoolGPA +20%	+2.49	+1.71	+1.98
HighSchoolGPA +10%	+2.05	+1.38	+1.65
SupportBin: 15K–20K	+2.64	+2.37	+1.03
SupportBin: > 20K	+2.51	+5.42	+0.96
TotalSupport +20%	+1.43	+1.18	+1.59
Matric Gender = Female	+1.92	+0.00–0.27	+1.90

TABLE V
NEGATIVE POLICY EFFECTS ON PREDICTED GRADUATION PROBABILITY

Policy Change	XGB Base	XGB Tuned	GB Tuned
Force = New Freshman	–8.16	–7.82	–9.32
Ethnicity = Hispanic/Latino	–5.25	–4.39	–5.54
Ethnicity = Black/African American	–3.98	–3.39	–3.98
Out of State Resident	–5.44	–4.27	–5.18
SupportBin < 5K	–2.20	–3.32	–0.16
HighSchoolGPA –20%	–3.03	–2.61	–2.30
TotalSupport –20%	–2.06	–1.77	–2.30
SAT/Eng/ALEKS –20%	–1.3 to –1.7	–1.2 to –1.7	–1.5 to –2.0

The three models produce nearly identical policy-sensitive effects, demonstrating that the underlying predictors exert stable influence regardless of algorithmic differences. The strongest positive effect occurs when assigning students the *New Transfer* status, indicating that transfer students bring clearer academic direction and stronger prior preparation.

International and Not Specified ethnicity categories lift predictions across all models, while Hispanic/Latino and Black/African American assignments reduce predicted outcomes, suggesting structural and institutional challenges that fall outside model behavior.

Financial support exerts one of the strongest influences. Students receiving more than $20,000 in aid or increasing TotalSupport by 20% consistently show higher predicted graduation probabilities. Academic improvements such as GPA increases yield moderate positive effects.

The largest negative impacts occur when assigning *New Freshman* status or reducing student academic preparation and financial support levels. These findings demonstrate that policy interventions related to academic readiness, financial stability, and student pathways can significantly alter predicted graduation outcomes.

E. Partial Dependence Analysis

The three models produce highly consistent partial dependence patterns that align with the results observed in the feature importance analysis. The graduation probability of students increases sharply when they move beyond the lowest English readiness band measured through *EngSCORE*, after which the effect stabilizes. This pattern demonstrates that improving basic English preparation yields substantial early gains, while further increases result in diminishing returns.

Financial support variables exhibit strong and nonlinear effects. Graduation probability increases substantially as TotalSupport rises, especially when support exceeds moderate levels. However, once funding reaches higher thresholds, the marginal benefits decrease and the prediction curve levels off. At the opposite end, students receiving minimal support show significantly lower predicted outcomes.

Academic preparation indicators display smooth and progressive relationships. Mathematics readiness measured through ALEKS and SAT Math generates steady increases throughout their score ranges. High school GPA and rank percentile demonstrate consistent positive effects, confirming that stronger academic backgrounds correspond to higher graduation likelihood.

A binary separation appears in the partial dependence plots for matriculation status: *New Transfer* students consistently achieve significantly higher predicted graduation probabilities than *New Freshmen* across all three models. This supports findings from previous research that well-prepared transfer students exhibit strong academic momentum.

AP-related variables show mixed patterns. *AP_max_score* produces a clear upward trend, while *AP_ct_computer* shows a decline at higher values, likely reflecting small or specialized student groups rather than true negative effects.

Overall, the partial dependence analysis confirms that financial support, transfer status, English readiness, and mathematics preparation remain the dominant factors shaping graduation probabilities across all models.

VI. Discussion

The research demonstrates that accurate graduation predictions can be generated using pre-enrollment and first-term student information when the dataset undergoes rigorous cleaning, de-duplication, and leakage protection. The strong performance of XGBoost and other tree-based models confirms that graduation outcomes arise from complex interactions among academic readiness, financial support, and student entry characteristics. The findings validate previous research on the importance of financial aid and academic preparation, while extending the evidence through operational, model-driven insights.

Across all interpretability methods feature importance, partial dependence, categorical gap analysis, and policy simulation the results reveal stable and consistent patterns. English and mathematics placement scores emerge as the strongest indicators of academic readiness, reflecting foundational skill levels that shape long-term academic performance. Financial support plays a similarly significant role: students receiving more than $20,000 in institutional aid experience the most substantial improvements in predicted graduation outcomes.

Transfer students consistently demonstrate stronger graduation prospects than first-time freshmen. This trend suggests that transfer pathways bring academic momentum and prior college experience that enhance completion likelihood. These results mirror national findings from NCES, PPIC, and CCRC, reinforcing the structural benefits of transfer pathways.

The broader implications for institutions include three primary areas where structural factors may create unequal academic outcomes. Although the predictive models identify where outcome differences occur, they do not specify underlying causes; this opens opportunities for targeted, equity-oriented interventions. Institutions may use these indicators to refine advising programs, redistribute financial support, or design academic readiness initiatives aligned with student needs.

The integration of the optimized model into a Power BI dashboard demonstrates the feasibility of deploying predictive analytics in operational decision-making contexts. The system supports real-time prediction generation, adjustable classification thresholds, and scenario-based policy simulations. These capabilities transform static research findings into an actionable decision-support tool for financial aid distribution, academic guidance, and strategic planning.

Overall, the combination of methodological rigor, interpretability techniques, and operational deployment underscores the value of predictive analytics when aligned with institutional policy needs. The results highlight how machine learning, when supported by transparent data preparation and robust evaluation methods, can enhance student success initiatives on a large scale.

VII. Conclusion

This research developed a complete predictive modeling system that integrates financial aid records, academic background information, placement test scores, and student characteristics to estimate undergraduate graduation success. The project established a comprehensive analytic foundation through rigorous data integration, preprocessing, deduplication, and leakage prevention. Among all evaluated models, the tuned XGBoost classifier demonstrated the strongest performance, achieving an AUC of 0.846 on the test dataset while also maintaining balanced precision and recall.

Interpretability analyses including feature importance, partial dependence, categorical gap evaluation, and policy simulation— revealed consistent and intuitive predictors across all modeling approaches. Academic readiness indicators, particularly English and mathematics placement scores, emerged as dominant factors. Financial support levels, especially institutional aid amounts exceeding $20,000, played a similarly influential role in shaping graduation outcomes. Transfer student status consistently improved predicted graduation likelihood, confirming national research on the strength of transfer pathways.

The study's findings reinforce existing literature demonstrating that early academic preparation and financial support are crucial determinants of student achievement. By quantifying these relationships within an operational predictive model, the research offers institutions actionable insights for policy development and resource allocation.

A significant contribution of this work is the system's operational deployment. The Power BI dashboard implementation translates predictive analytics into a real-time decision-support tool that allows users to run individualized predictions, adjust decision thresholds, and conduct scenario analysis. This enables institutions to make data-driven decisions related to financial aid distribution, academic advising, and strategic planning.

Overall, the research demonstrates that machine-learning techniques when paired with strong data preparation and interpretability frameworks can generate precise and practical insights into student success. The developed system provides a scalable foundation for future institutional analytics and supports continued advancements in student achievement research and deployment applications.

References

[1] S. M. Dynarski, "College Costs, Financial Aid, and Student Decisions," National Bureau of Economic Research, 2022. [Online]. Available: https://www.nber.org/system/files/working_papers/w30275/w30275.pdf

[2] L. Muraskin, J. Lee, and W. S. Swail, "Raising the Graduation Rates of Low-Income College Students," The Pell Institute, 2004. [Online]. Available: https://files.eric.ed.gov/fulltext/ED490856.pdf

[3] M. E. Dello Stritto, D. Skinner, N. Aguiar, and G. R. Underhill, "Financial Aid Matters," *Journal of Student Financial Aid*, 2025. [Online]. Available: https://ir.library.louisville.edu/cgi/viewcontent.cgi?article=1866&context=jsfa

[4] C. Dougherty, L. Mellor, and S. Jian, "The Relationship Between Advanced Placement and College Graduation," National Center for Educational Accountability, 2006. [Online]. Available: https://files.eric.ed.gov/fulltext/ED519365.pdf

[5] Challenge Success, “The Advanced Placement Program: Living Up to Its Promise?,” 2013. [Online]. Available: https://news.stanford.edu/__data/assets/pdf_file/0031/127498/ChallengeSuccess-AdvancedPlacement-WP.pdf

[6] National Center for Education Statistics (NCES), “Undergraduate Retention and Graduation Rates,” Condition of Education Report, 2023. [Online]. Available: https://nces.ed.gov/programs/coe/indicator/ctr

[7] U.S. Government Accountability Office (GAO), “Community Colleges: Students Who Transfer to Four-Year Schools Have Higher Graduation Rates,” GAO-17-331, 2017. [Online]. Available: https://www.gao.gov/products/gao-17-331

[8] Public Policy Institute of California (PPIC), “Successful Transfer Students Graduate at High Rates,” California’s Higher Education System Report, 2018. [Online]. Available: https://www.ppic.org/publication/californias-higher-education-system/

[9] Community College Research Center (CCRC), “What We Know About Transfer Students,” Teachers College, Columbia University, 2019. [Online]. Available: https://ccrc.tc.columbia.edu/media/k2/attachments/what-we-know-about-transfer-students.pdf