

WIP: The Impact of Financial Aid and Academic Pathways on Graduation

Chaturya Yarradoddi

CSEE Department

UMBC

Email: my81293@umbc.edu

Anupam Joshi

CSEE Department

UMBC

Email: joshi@umbc.edu

Abstract—Our work seeks to find out whether we can predict a student’s graduation in STEM majors using AI, specifically in Engineering and Computing, based on both academic factors known at admission time and financial factors that the institution can control. We create an integrated predictive modeling system that uses financial assistance in conjunction with academic preparation and enrollment methods to determine student success as measured by graduation. The dataset contains anonymized records of 13,694 students who matriculated at our university from 2013 to 2019. For each student, we have data on their high school GPA, placement test results, 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 de-duplication. We then trained machine learning models on these data.

Analysis of the models revealed that the amount of financial support, mathematics and English placement results, high school academic performance, and transfer student status were important factors in predicting graduation. Although these factors are intuitive, our analysis provides firmer, data-driven support and leads to a model that can predict graduation given these factors at admission time. It also provides information to advisors who can play with what-if scenarios, such as changing the amount of financial aid. The final model is implemented into a Power BI dashboard which enables users to run real-time predictions and adjust classification thresholds.

I. INTRODUCTION

Although Universities have always cared about graduation rates, there is now a greater focus on improving graduation rates because of the rising accountability standards being created by state legislatures and the federal government. The ability to predict student graduation success enables universities to direct their resources effectively while improving advising services and creating better intervention programs.

Financial aid is a key determinant of a student’s ability to stay in school and finish their degree, especially for students from low socio-economic status in STEM fields. Research findings show that students who receive more financial aid tend to achieve better graduation and retention outcomes [10], [1], [2], [3]. Academic readiness is another vital element that 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].

Although current research provides essential descriptive findings about graduation patterns, it does not provide a predictive system that could generate personalized probability estimates for potential or matriculated students. Our work creates a single analytical dataset to train machine learning models on student information before enrollment and their first academic term data. Our research identifies graduation predictors through machine learning, and also incorporates the model into an interactive Power BI dashboard. The system generates immediate predictions and threshold-based results and simulation tools that can help admissions and advising units 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. Our own prior research with a portion of the dataset we use here provides similar results[10].

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 our study match previous studies because the model consistently generated higher graduation probability estimates for transfer students.

Higher education institutions often use analytics to identify students who need academic or financial support. Most predictive models use restricted data sources, tend not to use AI/ML, depend on student behavior that has not yet occurred, and do not provide clear policy-relevant insights. The research builds upon previous work by uniting high school records, AP coursework, placement exam results, demographic information, and multiple years of financial aid data into a single modeling system that uses AI, specifically Decision Tree type machine learning models. Our research enhances both methodological precision and operational decision-making for student success through its implementation of strict data protection measures, interpretability and Power BI deployment.

III. DATA PREPARATION AND MODELING FRAMEWORK

This study integrates financial aid and academic preparation data from students admitted to UMBC. The financial aid data set contains 13,694 student records spanning six years, while the AP data set includes 58,246 exam records representing complete AP testing histories. These sources were combined into a single analytical database.

Financial support features were engineered to measure both the magnitude and structure of aid, including total support provided (*TotalSupport*), whether any support was provided (*Supported*), total support binned into groups (*SupportBin*, \$5K to >\$20K in \$5K steps), and *NeedStatus* based on FAFSA eligibility. This status had students that were Pell eligible labeled as High Need, students within twice the Pell eligibility limit labeled a moderate need, and others labeled as low need. AP exam data were processed to generate academic readiness metrics by grouping subjects into seven domains and extracting features such as exam counts, earned credits, score distributions, and STEM-focused indicators.

Standardization of student identifiers and duplicate resolution produced 13,694 unique student records after merging the datasets. Missing numeric values were imputed as zero, and categorical fields labeled as “Unknown.” The final dataset consists of 31 pre-enrollment and first-term features was used for both machine learning and policy evaluation.

The machine learning models used pre-enrollment information together with first-term data to create forecasts about student success, which was represented by a simple binary target variable indicating whether they graduated within six years.

Data were stratified into 70%/15%/15% groups to create training, validation and test sets. We trained a variety of decision tree based models, given that they provide explainable outcomes, unlike some deep learning approaches. The scikit-learn[11] pipelines built into Python were used to run all models with identical preprocessing steps. The analysis used median-imputation for numerical data which received standardization treatment while categorical data received the most common value for imputation and one-hot encoding. The RandomizedSearchCV optimization process built into scikit-learn for tree-based models was used to optimize model hyperparameters.

The evaluation system based its performance assessment on accuracy, recall, the F1 score (the harmonic mean of precision and recall), and the area under curve (AUC) metrics. However, AUC served as the main evaluation metric. We tested principal component analysis (PCA) based dimensionality reduction but it decreased both performance and model interpretability, so we removed it from the analysis.

IV. RESULTS

A. Model Performance

TABLE I
PERFORMANCE OF MACHINE LEARNING MODELS

Model	Accuracy	Precision	Recall	F1 Score	AUC
RF	0.7006	0.7275	0.8567	0.7868	0.7739
XGB	0.7116	0.7533	0.8224	0.7863	0.7636
GB	0.7061	0.7423	0.8338	0.7854	0.7631
GB_Tuned	0.7006	0.7467	0.8109	0.7775	0.7604
RF_Base	0.6950	0.7255	0.8481	0.7820	0.7577
LogReg	0.6969	0.7403	0.8166	0.7766	0.7402
LogReg_Tuned	0.6932	0.7414	0.8052	0.7720	0.7368
XGB_Base	0.6728	0.7228	0.7994	0.7592	0.7264
RF_PCA	0.6821	0.7164	0.8395	0.7731	0.7210
GB_PCA	0.6710	0.7292	0.7794	0.7535	0.7165
XGB_PCA	0.6470	0.7135	0.7564	0.7344	0.7042
DT	0.6724	0.7452	0.7481	0.7467	0.6618
DT_Tuned	0.6404	0.7248	0.7137	0.7192	0.6880
DT_PCA	0.6305	0.7105	0.7214	0.7159	0.6093

As shown in Table I, ensemble-based models consistently outperformed linear and single-tree approaches in predicting student graduation outcomes.

The Random Forest (RF) model achieved the AUC (0.7739) and the highest recall (0.8567). The Random Forest (RF) model can find graduating students while keeping false negatives low. The XGBoost (XGB) model attained an accuracy

(0.7116) and the highest precision (0.7533). The XGBoost (XGB) model also had an F1 score (0.7863). As such, the XGBoost (XGB) model balances, Recall and precision well. As can be seen with the results of not just XGB but Gradient Boosting (GB) and Random Forests, ensemble and boosting methods performed better overall.

We tried the simpler Logistic Regression models. These models gave lower AUC values. When given data modified by PCA based approaches, the models always performed worse, we suspect because dimensionality reduction took away some key decision features.

B. Feature Importance

The feature importance analysis indicates that financial support variables are the strongest predictors of graduation across all evaluated models. Gradient Boosting and XGBoost consistently assign high importance to *TotalSupport* and *SupportBin > 20K*, emphasizing the role of financial stability in supporting student persistence and completion.

Academic preparation variables also contribute substantially to graduation outcomes. Students with stronger prior performance, reflected by higher high school GPA, rank percentile, *CalScore*, *ALEKSScore*, standardized test scores, and Advanced Placement (AP) features, demonstrate higher predicted graduation probabilities. AP coursework performance and transferred credits further capture students' exposure to academic rigor prior to matriculation.

Enrollment-related factors, including ethnicity, part-time enrollment during the first semester, need-based financial aid status, and overall support levels, exhibit moderate importance. These variables represent secondary structural influences rather than direct measures of academic readiness. Across all models, the results consistently indicate that financial support combined with academic preparedness is the primary driver of graduation success. While this result is intuitive, our data provides empirical support. More importantly, we use these models we've trained to predict graduation outcomes for a new student given this background data.

C. Policy Simulation

The Policy simulation process allows us to see how outcomes change when a single variable is changed while others are held constant. This allows building admission strategies at an institution, understanding how financial aid for instance might affect graduation rates.

TABLE III
POSITIVE POLICY EFFECTS ON PREDICTED GRADUATION PROBABILITY

Policy Change	XGB Tuned	RF Tuned	GB Base
SupportBin = >20K	+2.55	+9.25	+0.96
Matric Status = New Transfer	+4.85	+1.61	+2.90
HighSchoolGPA +20%	+3.79	+1.57	+1.98
HighSchoolGPA +10%	+2.76	+1.02	+1.65
NeedStatus = No	+2.38	+1.76	+0.42
HS GPA Band = 4.0–4.49	+0.68	+0.68	+0.68
Supported = Yes	+0.52	+0.52	—
TotalSupport +20%	+0.51	+0.51	+1.59

TABLE II
FEATURE IMPORTANCE COMPARISON

Feature	XGB Tuned	RF Tuned	GB Base
Total Support	0.050	0.172	0.471
SupportBin > 20K	0.377	0.108	—
SupportBin 5K–10K	0.016	0.020	0.011
SupportBin < 5K	—	0.017	—
SupportBin No Support	—	0.017	—
HighSchool GPA	0.014	0.078	0.060
HighSchool Rank Percentile	0.015	0.028	0.027
CalScore	0.014	0.033	0.062
ALEKSScore	—	0.018	0.025
AlgSCORE	—	0.028	0.016
EngSCORE	—	0.016	0.018
SAT Math Score	—	—	0.017
SAT Reading/Writing Score	—	—	0.015
AP CRDS	0.021	0.044	0.059
AP Total Transfer Credits	0.014	0.039	0.019
AP Max Score	0.017	0.034	0.010
AP ct Language	0.014	—	0.010
AP ct Social	—	0.017	—
AP ct STEM	—	—	0.012
Sem1 FT/PT (PT)	0.019	0.015	0.009
Supported Yes	0.014	0.016	—
Need Status Yes	0.018	—	—
HS GPA Band 2.5–2.99	0.020	—	0.022
HS GPA Band 3.0–3.49	—	—	0.009
HS GPA Band 4.0–4.49	0.017	—	—

TABLE IV
NEGATIVE POLICY EFFECTS ON PREDICTED GRADUATION PROBABILITY

Policy Change	XGB Tuned	RF Tuned	GB Base
HS GPA Band = 2.5–2.99	-7.73	-6.57	-16.76
SupportBin = 5K–10K	-6.89	-6.89	-5.28
HighSchoolGPA -20%	-2.61	-2.30	-2.30
TotalSupport -20%	-1.77	-2.06	-2.30
SAT Math -20%	-1.14	-1.14	-1.70
SAT Reading/Writing -20%	-1.32	-1.32	-1.50
AlgSCORE -20%	-0.18	-0.18	-0.18
AP CRDS -20%	-0.06	-0.06	-0.06

The Random Forest Tuned, XGBoost Tuned, and Gradient Boosting Base models show similar though not identical policy-sensitive effects, demonstrating that the results around what variables can be tweaked to improve graduation rates remain mostly stable across different modeling techniques. Financial support emerges as the most effective positive policy lever, with *support > \$20K* and increases in total support producing substantial improvements in predicted graduation outcomes. Academic performance, including increases in high school GPA and placement into higher GPA bands, consistently generate positive effects on graduation as can be expected. The former can be used to increase graduation chances for a given student, and the latter can be used to shape

a class and predict graduation outcomes for a given class and given financial support strategy.

The strongest negative policy impacts arise from reduced academic readiness and lower financial support. The predicted graduation probabilities decline when students are in lower GPA bands, have lower standardized test scores, and less total financial aid. Across all three models, the consistent patterns highlight the central role of financial stability, academic preparedness, and student pathways in shaping graduation outcomes.

V. DISCUSSION AND CONCLUSION

Our research shows strong preliminary evidence that graduation rates can be effectively predicted by analyzing student academic performance and financial aid information at admission or matriculation. Using a variety of AI models, we show that Ensemble models consistently outperformed linear and single-tree approaches. This is likely because graduation outcomes arise from multiple interacting factors, including academic preparedness, financial assistance, and enrollment patterns. Random Forest achieved the highest recall and AUC values, while XGBoost provided the strongest accuracy–precision tradeoff.

Our analysis revealed consistent patterns in the factors affecting graduation across models. Our data shows that students receiving more than \$20K in financial aid and those demonstrating strong academic readiness through GPA and placement test scores achieve higher graduation outcomes, and upto a certain limit, graduation rates rise with increasing scholarship support. Policy simulations further showed that increases in financial support and academic preparedness lead to meaningful gains in predicted graduation probability, and decreases in these similarly lead to decreases in graduation rates.

The integration of the final models into a Power BI dashboard shown in Figure 1 enables real-time prediction and scenario-based policy evaluation for a given student, transforming research findings into an operational decision-support system. For a given student, an admissions officer can see how much a given financial aid package will change the gradtion results. Overall, the study illustrates how interpretable machine learning models can support data-driven institutional policies and improve student success outcomes.

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>
- [10] Solipuram, Akhila and Joshi, Anupam and Perman, Jay, “WIP: The effect of scholarships and community on graduation rates for STEM majors,” 2024 IEEE Frontiers in Education Conference (FIE), pp 1–5, 2024
- [11] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, 12, pp2825–2830, 2011

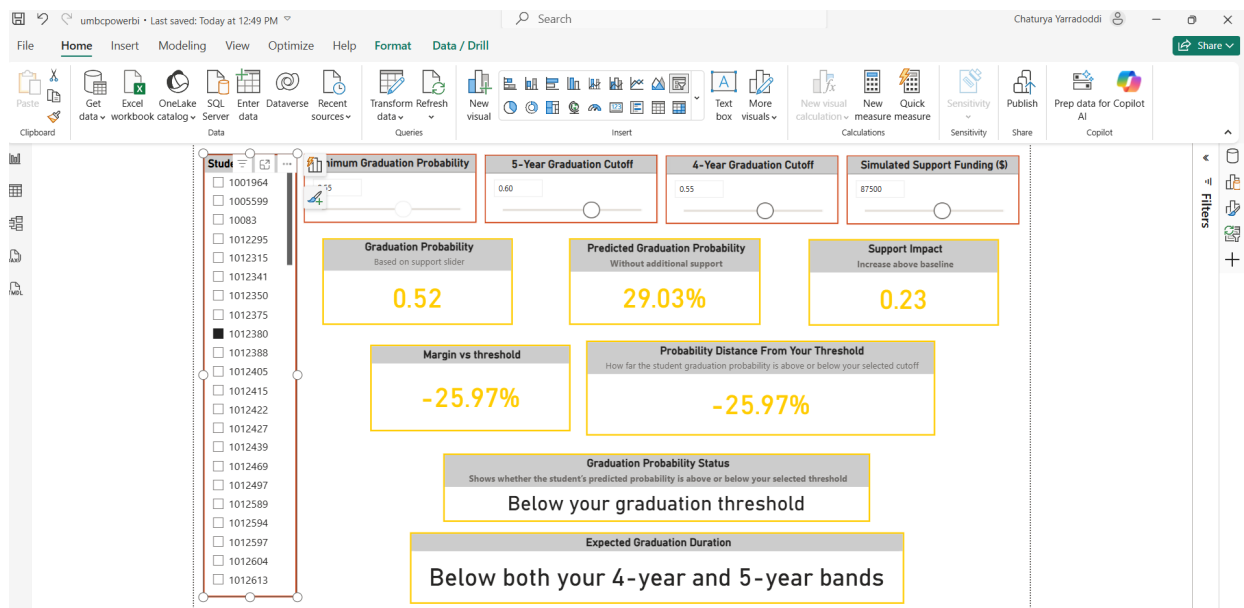


Fig. 1. Power BI dashboard for real-time graduation prediction and policy simulation