

Gateway or Closed Door?: Exploring the Effect of Gateway Courses on College Success

Matthew Fikes

DSC 680, Fall 2021

SUMMARY

This study explores connections between student performance in gateway courses to later academic success in the setting of a community college. Gateway courses are defined as high-enrollment (>100/year) and high failure rate (>30%) and are traditionally the first college level for-credit course a student takes in their program. The goal is to see if the number of courses that result in D and F letter grades in addition to Incomplete and Withdrawal can be used to predict semester-to-semester retention in new community college students.

The strongest predictor in an ordinary linear squares model is shown to be not the rate of D/F/I/W grades in gateway courses, but rates of those grades overall. An OLS model showed some success but predictions were too low to be useful. The decision tree gave better results, predicting correctly nearly 80% of the time. While the predictive power of DFIW grades is strong, it is not limited to gateway course success as strongly as it is linked to overall success.

INTRODUCTION

BACKGROUND

When exploring ways to improve academic success in community colleges, gateway courses are a good place to start. Gateway courses are generally considered the first credit-bearing class a student will take within their program of study. For this project I will be using the definition that comes from Achieving the Dream, (*About us: Achieving the Dream*, n.d.). This looks for courses with high enrollment and high failure rates, with a disparity in success between students with and without low income. Our institution came up with figures of >100 enrollments in an academic year for a given cohort and a success rate <70%. A study on gateway course completion by Gordon Flanders found that first-time full-time freshmen were more likely to persist if they successfully completed a gateway course compared to other groups (Flanders, G.R., 2015.)

PROBLEM STATEMENT

Is gateway course performance an indicator of academic success?

SCOPE

The project covers student cohorts for Fall semesters in 2016 and 2017. Allowing for more than one year of data should help inform the model and having the most recent cohort data several years back allows for follow-up analysis on outcomes at 150% of the time estimated to earn a degree. The project will not look at whether instructors were full-time or adjunct or the features of the specific course sections.

METHODS

DATA SOURCES

This data comes from the student data system of a community college that utilizes Ellucian Banner (Banner ERP System for Universities, n.d.). Any personally-identifiable information (PII) has been removed or obscured for compliance with federal privacy laws.

DATA PREPARATION

For processing of the data, the features needed to be created from the student data system records. The total number of gateway and total classes were counted for the first semester of school for each student and listed in **GatewayCount** and **CourseCount** features. **FS_RET** and **FF_RET** are target variables that indicate whether a student was retained from Fall-to-Spring and Fall-to-Fall respectively. **GPA** is the final term grade point average on a 4.0 scale. The percent of classes taken with letter grades D, F, I, or W were calculated for all classes and gateway classes for the creation of features **DFIW** and **DFIW_Gateway**.

Data was split into testing and training data with sklearn with a 60/40 split.

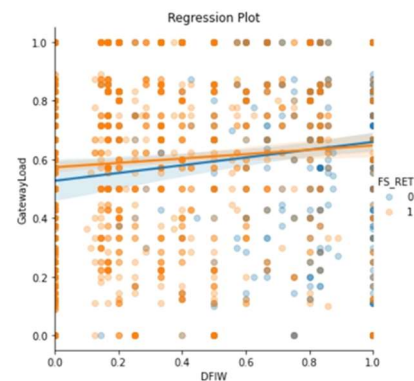
MODELING

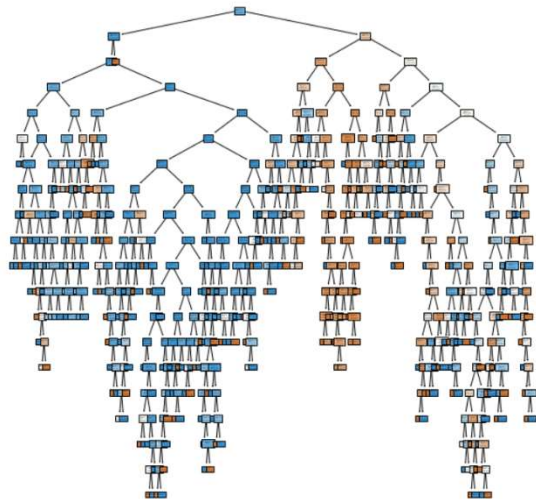
I chose to use ordinary least squares regression to test for linear relationships. In case linear relationships are not evident, I will be using a decision tree classifier to attempt predicting retained students.

RESULTS

The results of the linear model do not seem to indicate the best solution for this problem. Summary output shows an R-squared of only 0.338, indicating a weak, if any, affect. A plot of DFIW rate against GatewayLoad shows that the difference in retention is found when

DFIW varies more so than when the percent of gateway courses taken does. Similarly, the summary output indicated that DFIW had a larger coefficient than DFIW_Gateway.





The decision tree model fared better than the linear regression. Using sklearn metrics to check the model predictions resulted in an approximate 80% accuracy. The resulting tree is large and may benefit from pruning, but the 80% accuracy was without any tuning of hyperparameters.

The results seem to indicate that gateway course performance isn't necessarily an indicator of student retention, but overall performance is.

ASSUMPTIONS

These students are all assumed to be first-time at any school, as reported on their application. Gateway course definitions are assumed to remain consistent since they were selected. There are no duplicate students and the students selected should be from a diverse set of demographics that were not controlled for in this study.

LIMITATIONS/CHALLENGES

The dataset is only covering one community college and so any results may not apply to 4-year schools. The collected cohort semesters were all prior to COVID so it may be worth doing a second comparison after enough data is available on outcomes post-COVID.

NEXT STEPS

I would want to broaden the dataset with additional semesters of data. That will make the model training more robust and allow better understanding of trends over time.

The decision model was trained with no tuning of the hyperparameters and may benefit from some fine-tuning.

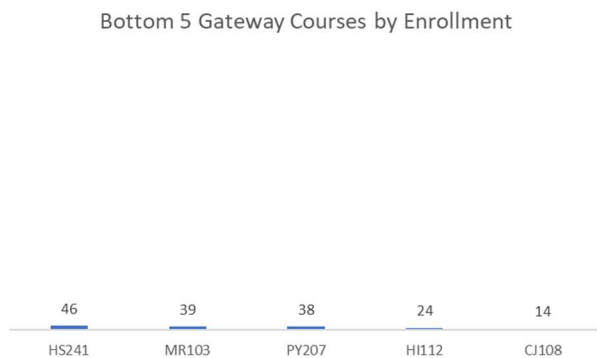
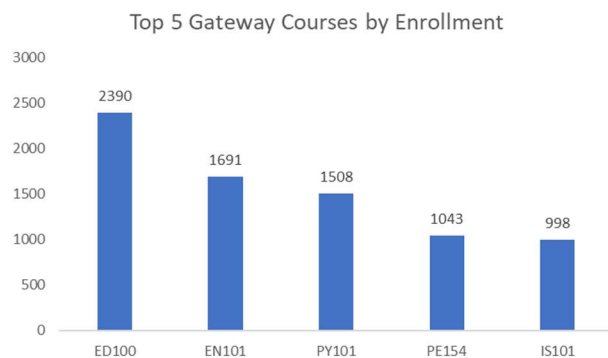
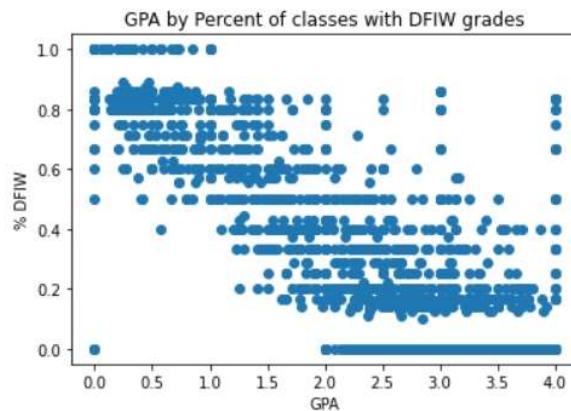
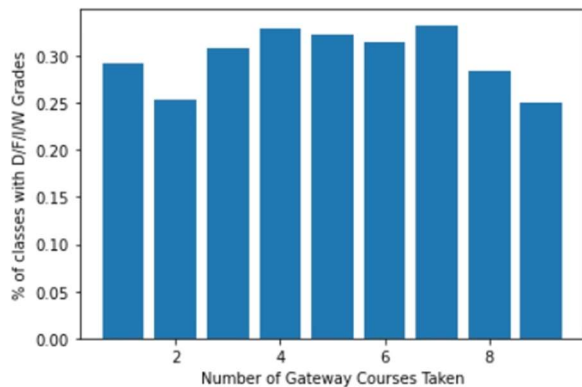
I would like to go back through and test some parameters with GridSearchCV to see if results could improve even further.

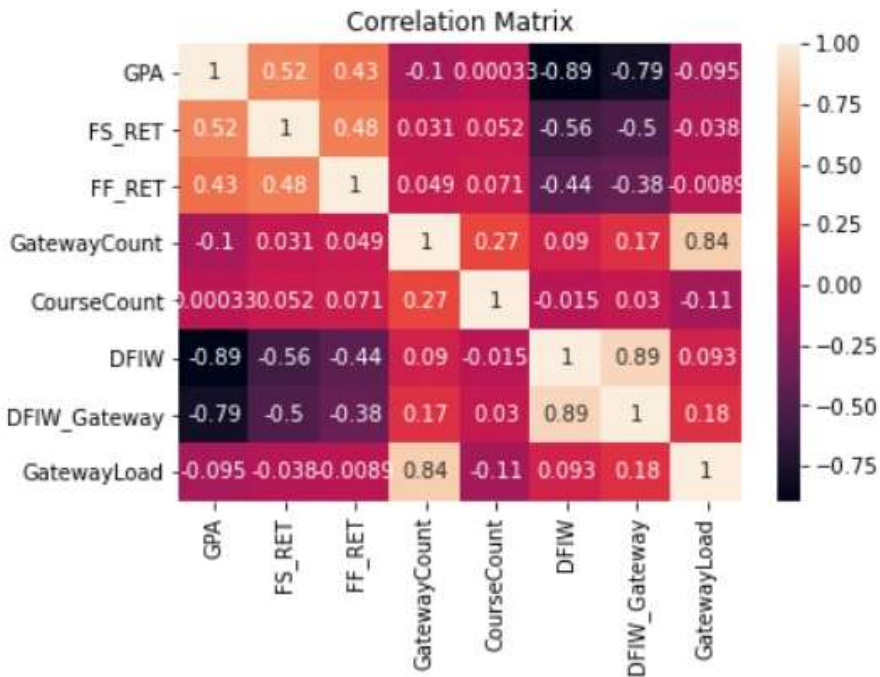
Any additional data collected from other terms would help prevent overfitting of the decision tree, as the current dataset is not large.

Exploring how performance in certain classes like Math or Chemistry compared to Psychology or Physical Education pre-requisites could make feature-selection more targeted.

There is also merit at investigating how faculty status (full time vs adjunct) affects retention. A 2020 study by Florence Ran and Jasmine Sanders found that community colleges have grown in overall percentage of adjunct instructors. While success rates were comparable, subsequent enrollment deviated -7% from the sample mean. This could have an effect on retention numbers, benefiting students and the school.

Appendix:





OLS Regression Results

Dep. Variable:	FS_RET	R-squared:	0.338
Model:	OLS	Adj. R-squared:	0.335
Method:	Least Squares	F-statistic:	131.0
Date:	Thu, 21 Oct 2021	Prob (F-statistic):	4.50e-134
Time:	03:53:55	Log-Likelihood:	-515.51
No. Observations:	1547	AIC:	1045.
Df Residuals:	1540	BIC:	1082.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	0.9800	0.063	15.604	0.000	0.857	1.103
GPA	0.0273	0.014	1.920	0.055	-0.001	0.055
GatewayCount	0.0727	0.011	6.704	0.000	0.051	0.094
CourseCount	-0.0200	0.005	-3.706	0.000	-0.031	-0.009
DFIW	-0.5729	0.070	-8.236	0.000	-0.709	-0.436
DFIW_Gateway	0.0024	0.051	0.047	0.962	-0.098	0.103
GatewayLoad	-0.3975	0.070	-5.690	0.000	-0.535	-0.260

Omnibus:	146.157	Durbin-Watson:	2.095
Prob(Omnibus):	0.000	Jarque-Bera (JB):	204.174
Skew:	-0.739	Prob(JB):	4.62e-45
Kurtosis:	3.990	Cond. No.	89.1

References:

Banner ERP system for Universities. Ellucian. (n.d.). Retrieved October 11, 2021, from <https://www.ellucian.com/solutions/ellucian-banner>.

About us: Achieving the Dream. (n.d.). Retrieved October 14, 2021, from <https://www.achievingthedream.org/about-us-0>.

Flanders, G. R. (2015). The effect of Gateway course completion on freshman college student retention. *Journal of College Student Retention: Research, Theory & Practice*, 19(1), 2–24. <https://doi.org/10.1177/1521025115611396>