Projecting the Graduates – and Dropouts – for Early Intervention

Darryl Patteson

Beatriz

- Single, First-Generation Student
- Advertising & Marketing Management
- Her tuition is up-to-date
- She has debt
- Not on scholarship
- 5 Units Approved Semester 1
- 5 Units Approved Semester 2

We know all of this, but will Beatriz go on to graduate?

Rodrigo

- Single, First-Generation Student
- Tourism Major
- His tuition is up-to-date
- He has debt
- Not on scholarship
- 6 Units Approved Semester 1
- 2 Units Approved Semester 2

We know all of this, but will Rodrigo go on to graduate?

Why Does It Matter?

If we can project dropouts, we can support faster and more directly

Targeted support should lead to more graduates

The big question: Who do we target?

So, What Data Do We Have?

- Detailed information on 4424 current and former students
- 35 factors attached to each students
  - Family's career and education information
  - Application status and age
  - Financial information for student and globally
  - Academic information for all students, including career path
- Target Feature: Graduate, Dropout or Enrolled Currently
  - 2209 Graduates
  - 1421 Dropouts
  - 794 Currently Enrolled

Findings: 5 Top Factors for Graduation

- Curricular Units 2nd Semester (approved)
- Curricular Units 1st Semester (approved)
- Tuition Fees Up to Date
- Curricular Units 2nd Semester (grade)
- Scholarship Holder

Factor Group #1: Freshman Performance

- Semester 1 & Semester 2 Units Approved
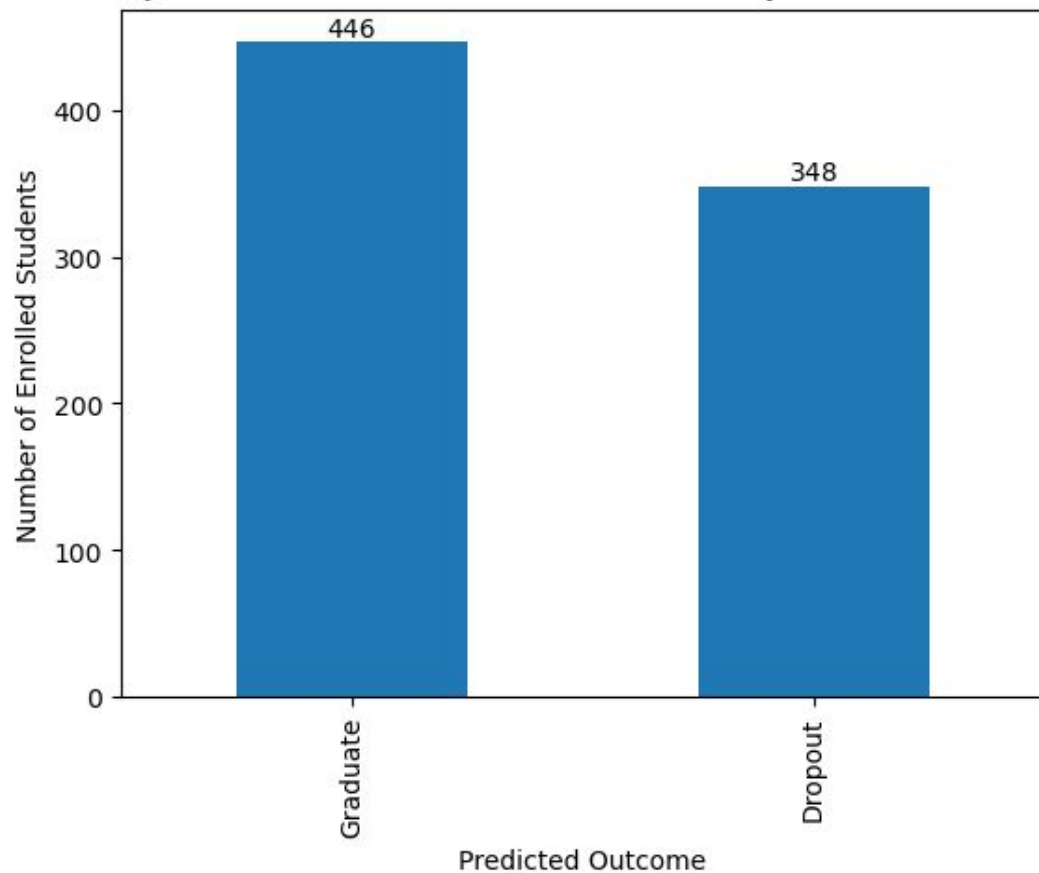- Semester 2 Grades

Factor Group #2: Financial Situation

- Tuition Fees Up to Date
- Scholarship Holder

Projected Graduation Outcome for Currently Enrolled Students
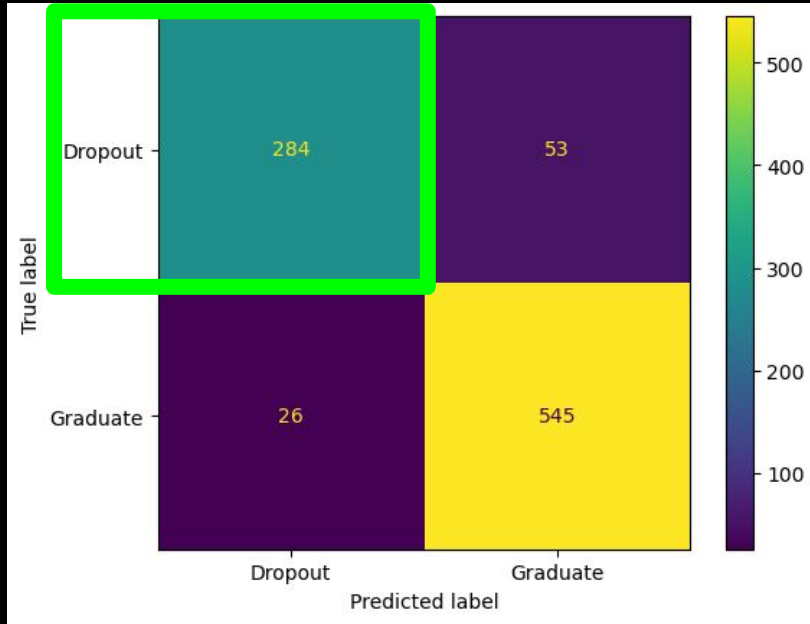
Model Generation

- Graduate & Dropout data was standardized and binarized
- Five models were developed and tested:
    - Logistic Regression
    - Random Forest
    - K-Nearest Neighbors
    - Decision Tree with Principal Component Analysis
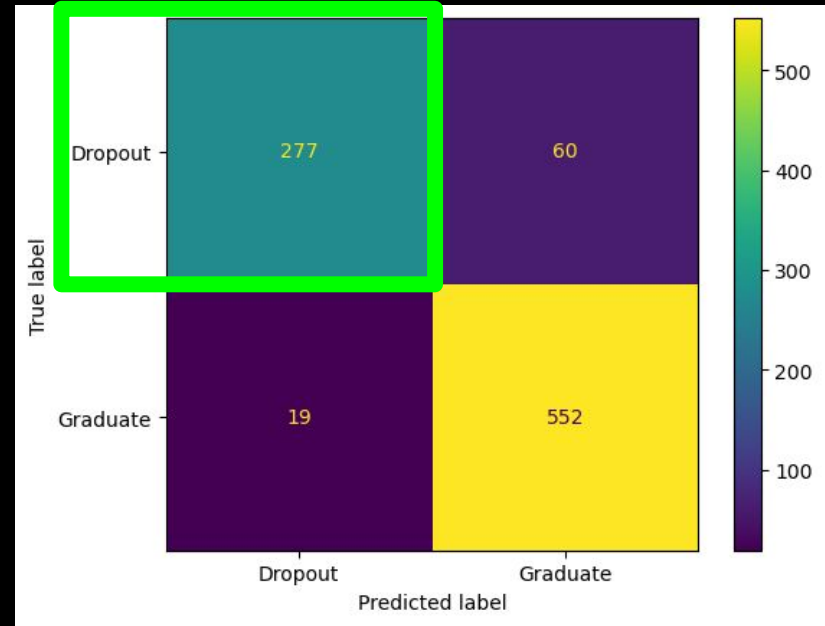   - Support Vector Machine (SVM)

Model Performance

Model	Dropout Precision	Dropout Recall	Dropout F1	Overall Accuracy
Logistic Regression	0.92	0.84	0.88	91%
Random Forest	0.92	0.81	0.86	90%
KNN	0.87	0.64	0.74	83%
PCA w/ Decision Tree	0.72	0.74	0.73	80%
SVM	0.94	0.82	0.88	91%

Model Selection: Confusion Matrix
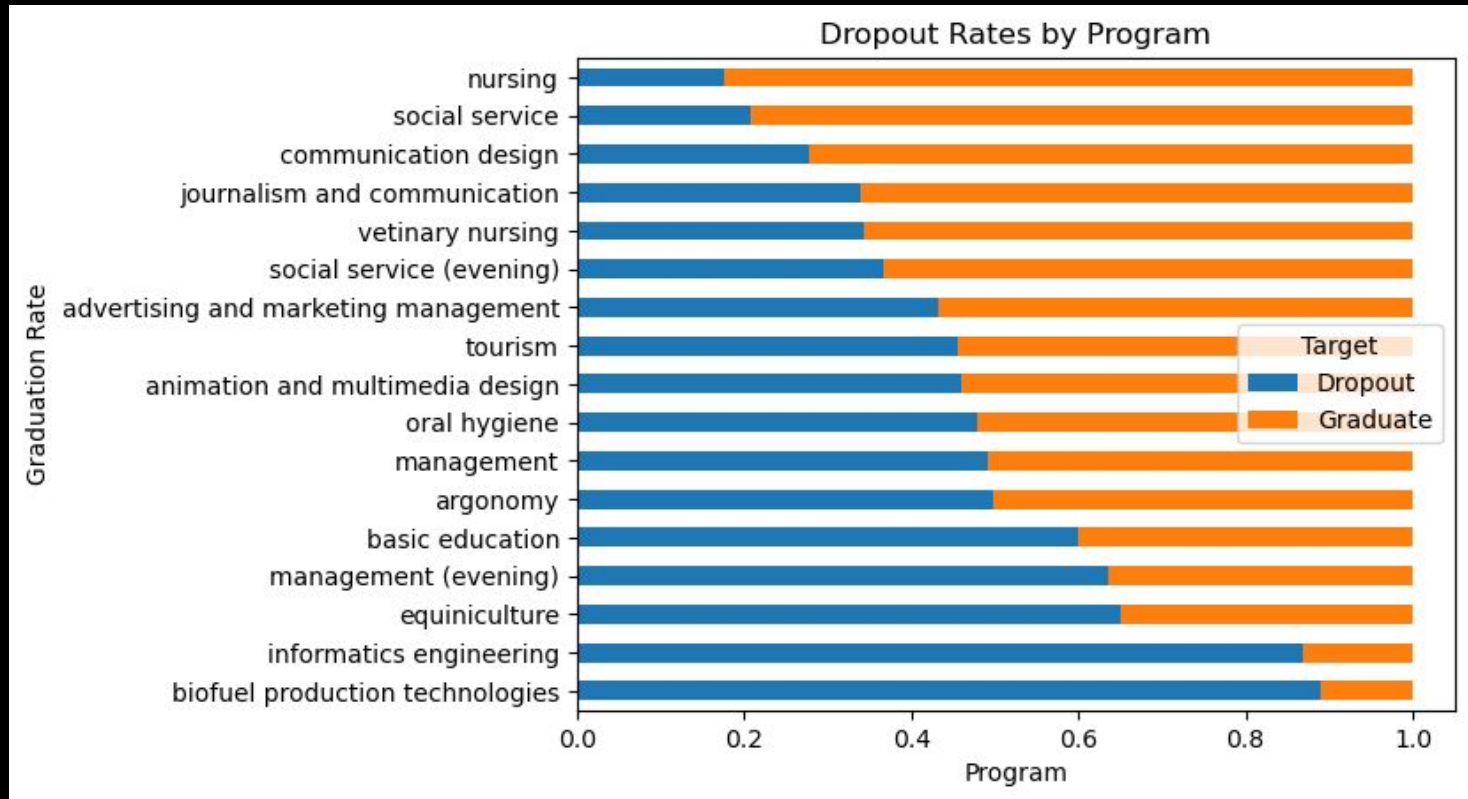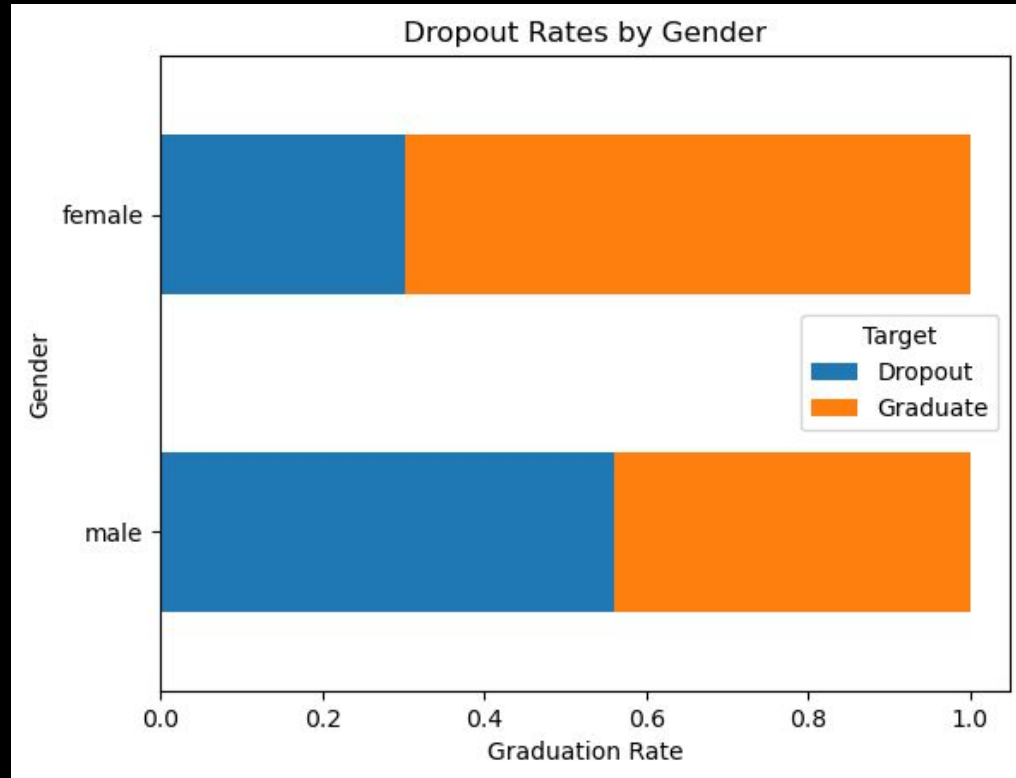
Uses of the Model

- Projected onto all currently enrolled
- We have all of their data and their projection in one place
- Use of the projection
  - Increase financial supports
  - Increase mentoring - academic Choices
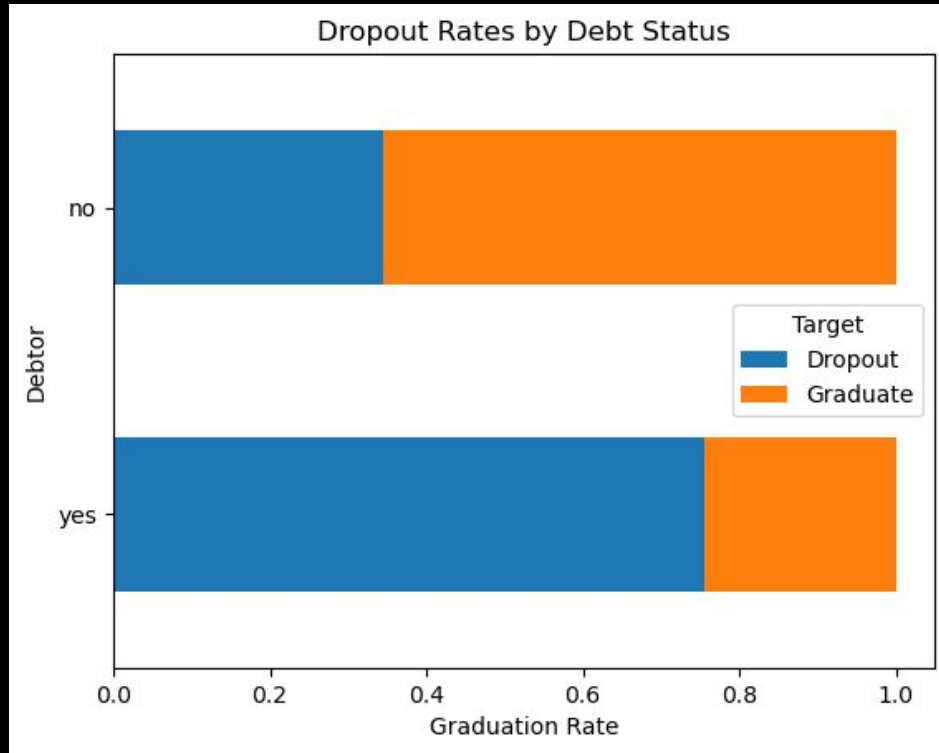 - Increase academic support for those missing credits

Further Examination: Program

Dropout Rates by Program

Further Examination: Gender

Further Examination: Debt

Beatriz

- Single, First-Generation Student
- Advertising & Marketing Management
- Her tuition is up-to-date
- She has debt
- Not on scholarship
- 5 Units Approved Semester 1
- 5 Units Approved Semester 2

Projection: Graduate

Rodrigo

- Single, First-Generation Student
- Tourism
- His tuition is up-to-date
- He has debt
- Not on scholarship
- 6 Units Approved Semester 1
- 2 Units Approved Semester 2

Projection: Dropout

What Can We Do For Rodrigo?

- Financial Support
  - Tuition is up to date, but will this stay the same?
  - Scholarship Opportunities
  - Work Study or Other On-Campus Option to Earn Money
- Academic Mentoring
  - Opportunities for mentoring
  - Study Groups
  - Personalized Office Hours
- Academic Counseling
  - Is the field of study the the right fit?
  - Summer credit hours to get back on track

Conclusion

- The model and data are a beginning points
- Using the model to work with potential dropouts should decrease dropout rates
- Incoming students can now be identified on the financial lines to know where most support is needed
- There are three areas to research more that could also decrease dropouts