BrainStation

Capstone Sprint 3

At-Risk Students - Support for Success

Bobbie Razo 2024.05.05

Contents

Escalating College Dropout Rates: A Growing Concern

With historical student data, can student drop out actions be predicted by identifying critical indicators?

We hope to use Data Science To Recognize Clues And Patterns That Precede College Student Dropouts.

Early Identification of At-Risk Students → **Early Intervention** → **Student Succeeds**

Implement Student Support Strategies

⬆ Student Retention Rates

⬇ Unemployment Rates (higher for College dropouts)

⬆ Earnings of Students as bachelor degree graduates

Data & Preprocessing

Collection of Student Enrollment 37 Features

Includes: - Academic path (curriculum credits, course, grade, schedule)
- Demographics (age, marital status, gender)
- Social-Economic Factors (parent info, financial)

Binning (Bucketing) of 8 High Cardinal Categorical Columns

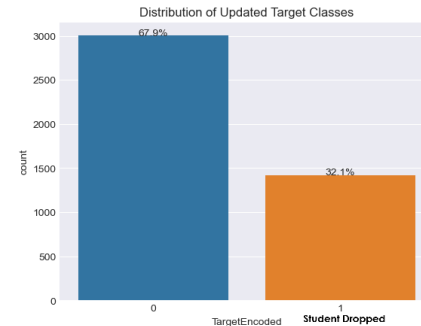Using: 1 – Domain Specific Knowledge:
ex. Educational Qualification

Qualification (Student & Parents): 34 Distinct Values	New (Binned) Qualification Values: 9 Distinct Values
1 - Secondary Education - 12th Year of Schooling or Equivalent	Unknown
2 - Higher Education - Bachelor's Degree	Nonliterate
3 - Higher Education - Degree	Elementary
4 - Higher Education - Master's	Middle School Incomplete
5 - Higher Education - Doctorate	Middle School
6 - Frequency of Higher Education	Secondary Incomplete
9 - 12th Year of Schooling - Not Completed	Secondary
10 - 11th Year of Schooling - Not Completed	Higher
11 - 7th Year (Old)	Vocational
12 - Other - 11th Year of Schooling	
13 - 2nd year complementary high school course	

2 – Manual by Frequency
ex. Nationality is binarized: 1 out of 21 values is 97% (Portuguese - 1)
other 20 is put into *Other* – 0

Target

- Binarized from 3-class Category to Dropped (1) vs. Not Dropped (0)
- Distribution is Imbalanced: Up Sampled the Minority Class (30% Dropped)

Distribution of Updated Target Classes

EDA Insights

- **Feature Correlation with Target**

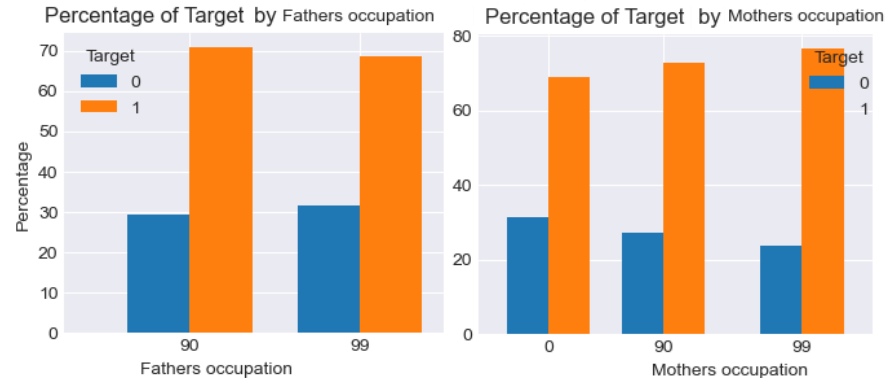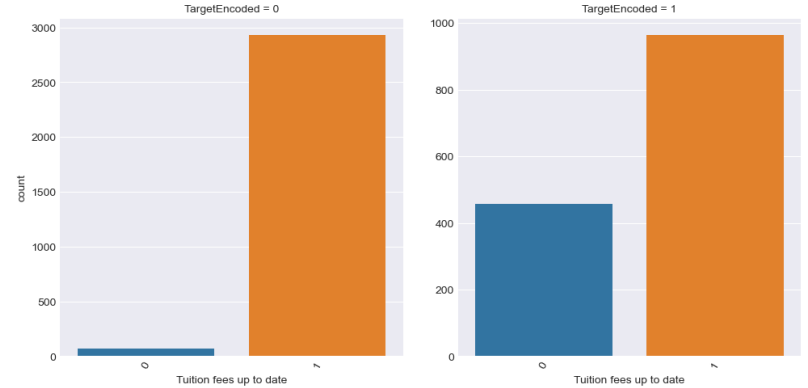 - **Ex. Tuition Fees Up To Date** – higher for Dropped.

- **Higher Dropped Percentage Rate for Certain Features Values**

  - Course, Marital Status, Age and Occupations

  - Anomaly OR a very small population causes the high percentage.

  Ex. Parent Occupation: 90 - Other Situation

  99 - (blank)

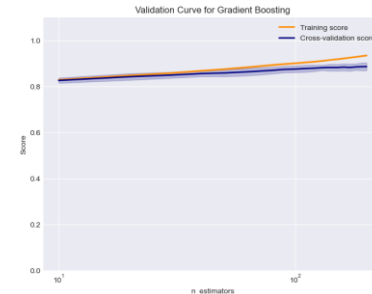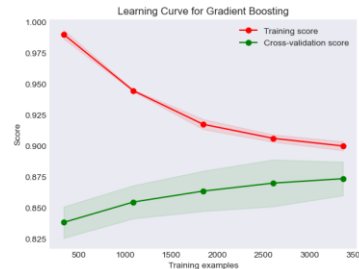 or Age of Enrollment tends to have higher drop rate with higher age.

Modeling Comparison

Models used:
- Logistic Regression
- PCA Reduction
- KNN
- Random Forest
- Decision Tree
- Gradient Boosting Classifier

Evaluations Used:
- Accuracy Score
- Classification
- Cross-Validation
- ROC Curve
- Learning Curve
- Validation Curve

Model	Accuracy	Training	Mean	Precision	Recall	F1	ROC
Random Forest estimator=100	93.84	100.0	92.96	92.306	94.02	93.15	0.98
Decision Tree	89.34	100.0	86.988	84.126	91.822	87.79	0.90
KNN k=1 with 98% PCA k=5	89.28	100.0	77.886	84.702	68.616	75.798	0.87
Gradient Boosting Classifier	87.45	89.70	87.512	89.35	85.702	87.482	0.94
Baseline - Logistic Regression | 25% test data set	86.88	87.61	86.39	88.142	84.376	86.218	0.93
Baseline with 98% PCA	86.61	86.58	86.124	88.006	83.936	85.916	0.92
Baseline with 95% PCA	86.06	87.85	87.37	87.692	70.002	77.83	0.92
Baseline with 93% PCA	85.81	85.85	85.236	88.052	81.868	84.836	0.92
KNN k=5 with 98% PCA k=5	79.96	87.94	79.372	80.856	77.462	79.106	0.89
KNN k=5 with 95% PCA k=5	79.96	88.23	79.308	80.99	77.066	78.974	0.89
KNN k=5 with 93% PCA k=5	79.82	87.92	79.218	80.88	77.022	78.896	0.88

Learning Curve for Gradient Boosting

Validation Curve for Gradient Boosting

Potential Product

A Dashboard Report for Colleges

Student Risk Report Analysis

Semester	Academic Level	Student ID	Student Name	GPA	Enrolled Credits	Cummulative Credits	Tution Info	Scholarship Info	Factors
Fall 2023									
	Early 3rd Year	7456289031	R********	3.2	15	75	Up To Date	500	Single, parent
	Late 1st Year	2097152980	S********	2.8	15	30	Up To Date	1000	Single
	Senior	3857160492	T********	3.7	16	116	Up To Date	1500	Single
		5619274038	H********	2.6	12	132	Up To Date	0	Divorced, full time work, parent
	Sophomore	1845729048	M********	2.3	12	72	Up To Date	0	Divorced, full time work
Spring 2024									
	Early 1st Year	7890699699	J********	3.1	12	4	Up To Date	0	Married, Full time Work
	Early 2nd Year	1048576320	B********	3.5	18	48	Up To Date	2500	Single, Part time work
	Early 4th Year	6724358910	K********	3.8	15	105	Payment Plan	1000	Single
	Junior	3081756924	D********	3	18	98	Up To Date	2000	Single
		6394710258	L********	3.9	20	90	Up To Date	3000	Married, parent
	Late 2nd Year	9203874561	A********	2.5	9	60	Payment Plan	0	Married, part time work

Monday, May 6, 2024

Page 1 of 1

Thank you

References:

- Learnopoly College Dropout Rate Statistics

- Educationdata.org College Dropout Rates

- Data Source: Predict Students' Dropout and Academic Success