Brainstation

Capstone Sprint 2

At-Risk Students - Support for Success

Bobbie Razo 2024.03.12

Contents

Escalating College Dropout Rates: A Growing Concern

With historical student data, can student withdrawal 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

Implement Student Support Strategies

Student Succeeds

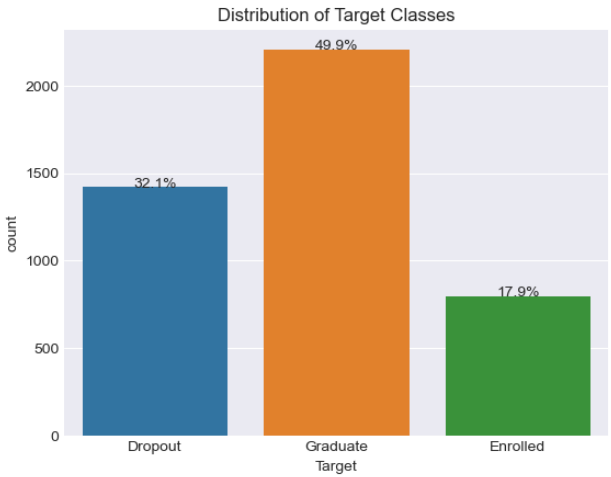
Increased Retention Rates

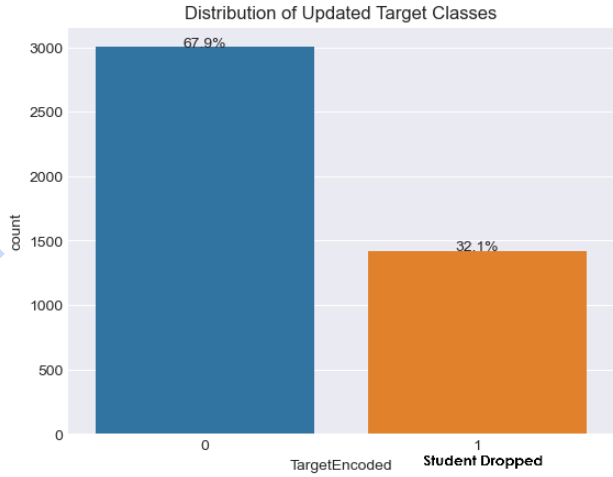
Data

- **Features:** Age | Gender | Country GDP | Country Unemployment Rate | Country Inflation Rate | Day/Night Schedule | Course | Marital Status | Nationality | International Admission Grade | Qualifications (including Parents) | Parents Occupation | Scholarship Info | Tuition Info | Curricular Units | Target – Dropout / Not Dropout

- Collection of student enrollment – academic path, demographics, and social-economic factors.

- Binarize the Target Variable: Dropped or Not

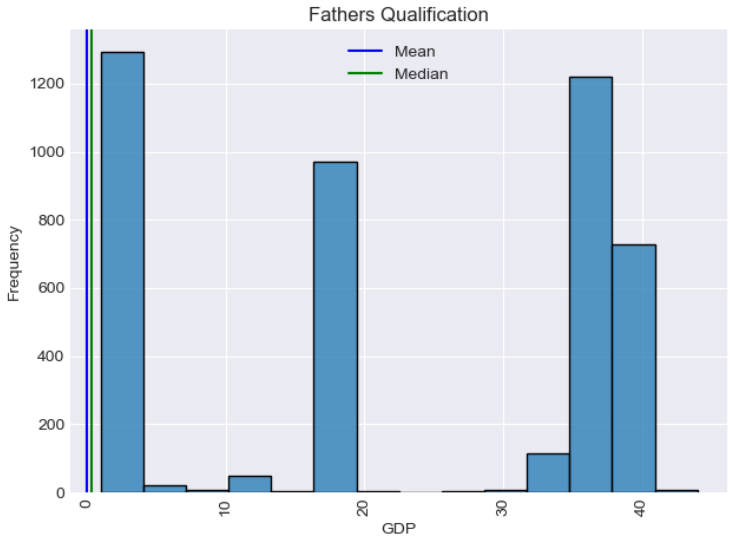


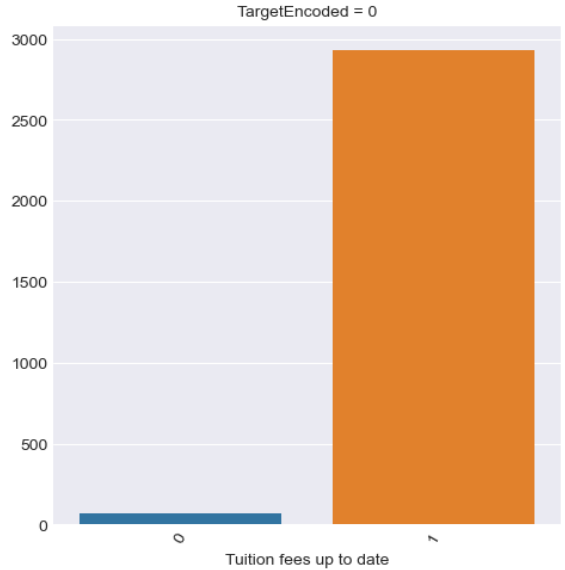


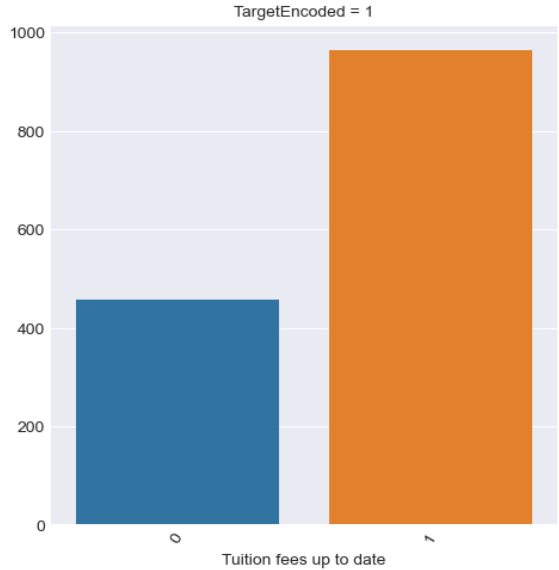
EDA

- **Numerous features with high cardinality**
  - **One-Hot Encoded**
  - **244 features for X data**

- Some features show a correlation with the Target: ex. Tuition fees up to date. This shows higher numbers of student struggling with tuition payments that result in dropping out.







Basic Modeling

- Initial modeling using Logistic Regression gives a somewhat good accuracy

Accuracy: 84.75
Train Accuracy: 87.88
Test Accuracy: 84.75

Classification Report:

	Precision	Recall	F1-Score	Support
0	0.85	0.93	0.89	569
1	0.85	0.70	0.77	316
Accuracy			0.85	885
Macro Avg	0.85	0.81	0.83	885
Weighted Avg	0.85	0.85	0.84	885

- PCA with ***.95*** & .90 component

	Train Score	Test Score
Base Logistic	87.88	84.75
PCA 95%	89.46	85.08
PCA 90%	89.09	84.41

Next Steps

1. **Further Feature Engineering:**

   - Binning of categorical features with high cardinality using domain knowledge and .map or .cut

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

New (Binned) Qualification Values: 9 Distinct Values
Unknown
Nonliterate
Elementary
Middle School Incomplete
Middle School
Secondary Incomplete
Secondary
Higher
Vocational

2. **Use of Up/Down Sampling for the unbalanced Features: Target variable**
3. **Assess using the SWOT Analysis (Strength, Weakness, Opportunities and Threats)**
4. **Strategize & Implement further Modeling: Decision Tree, SVM, Hyperparameters and Pipeline**

Thank you

References:

- Learnopoly College Dropout Rate Statistics
- Educationdata.org College Dropout Rates
- Data Source: Predict Students' Dropout and Academic Success