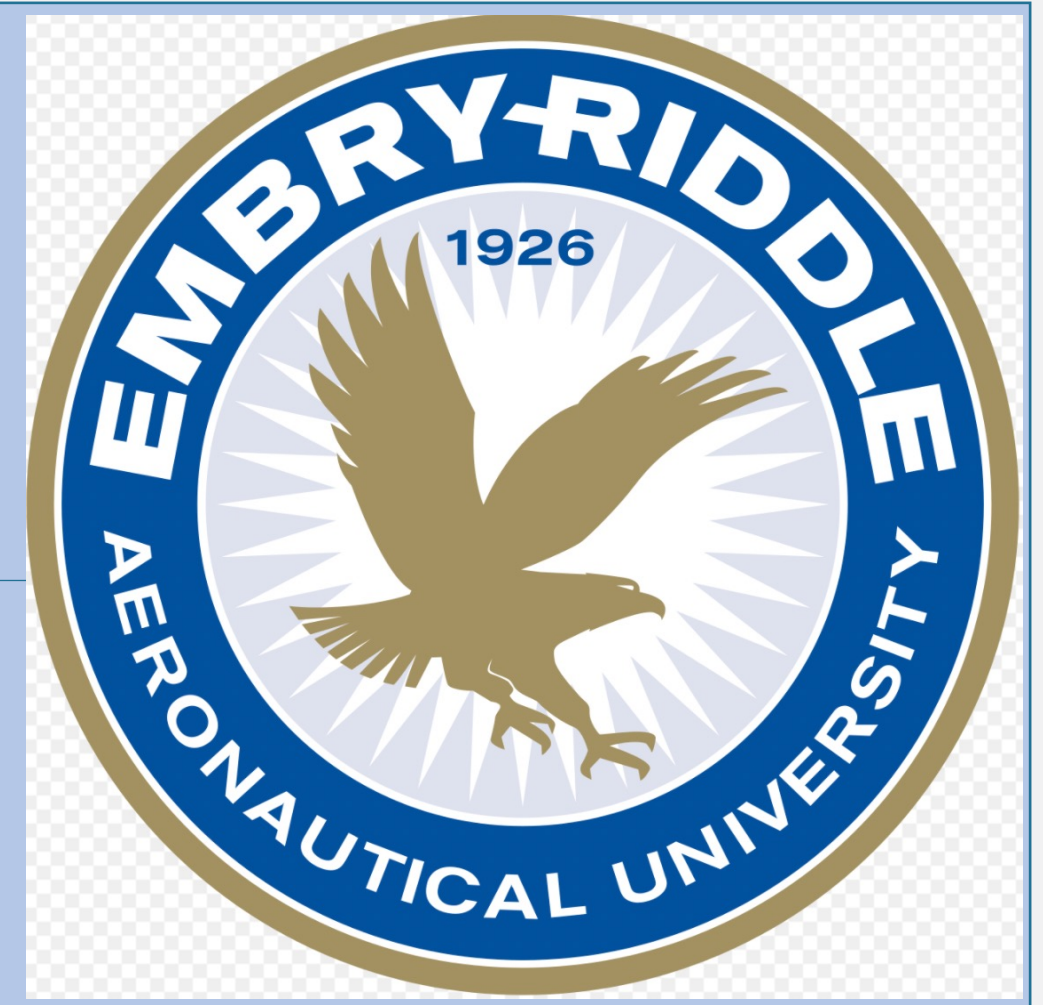# Predicting Undergraduate Students Major Switching

Elif Cankaya, Kyle Garber, Hong Liu

Background

Undergraduate students make an important decision when they select a major to enter a field at the start of their academic education, and eventually a career. After declaring major, however, students may have to change majors to move forward and earn a degree. Switching is quite common among students in science, technology, engineering, and mathematics (STEM) and major change is related to attrition rates in higher education.

Study Purpose

Given the importance of increasing persistence in STEM fields and of timely graduation, this study aimed to investigate factors that may affect undergraduate students' major switching.

Methodology

This study uses student data from ERAU, which was provided in excel. The dataset was cleaned and only key attributes that have the possibility to influence whether a student switches majors were extracted.
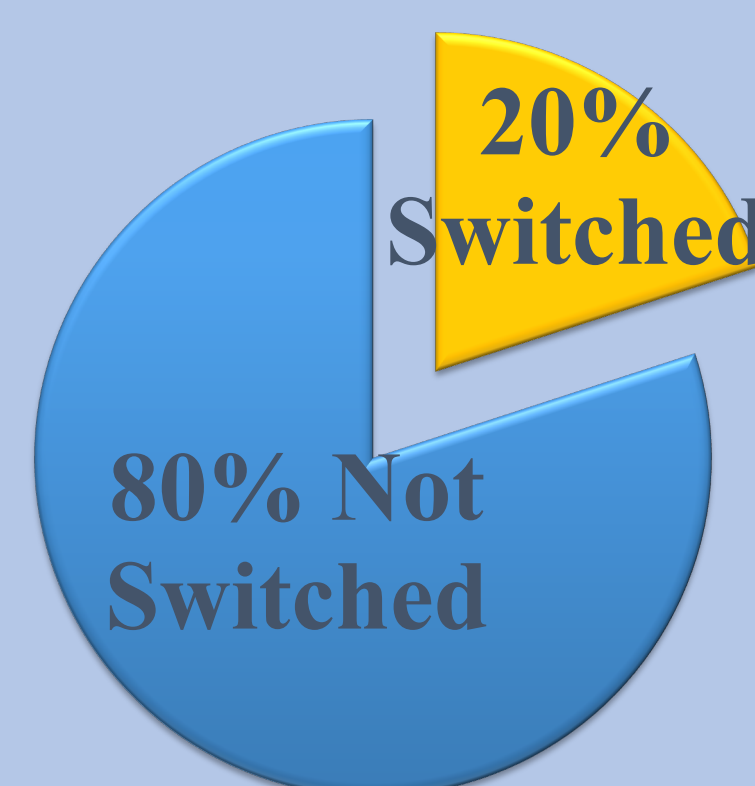
Sample

The sample includes a total of 12,441 undergraduate students. While 2500 students switched major, 9, 941 students did not.

Major Switched or Not

20% Switched
80% Not Switched

Data Processing

Raw Student Data
ERAU provided excel files

Data Cleaning
Eliminate extraneous data

Data Transformation
Key attributes were selected

Exploratory Analysis

Visualization

Run Models
Compare results
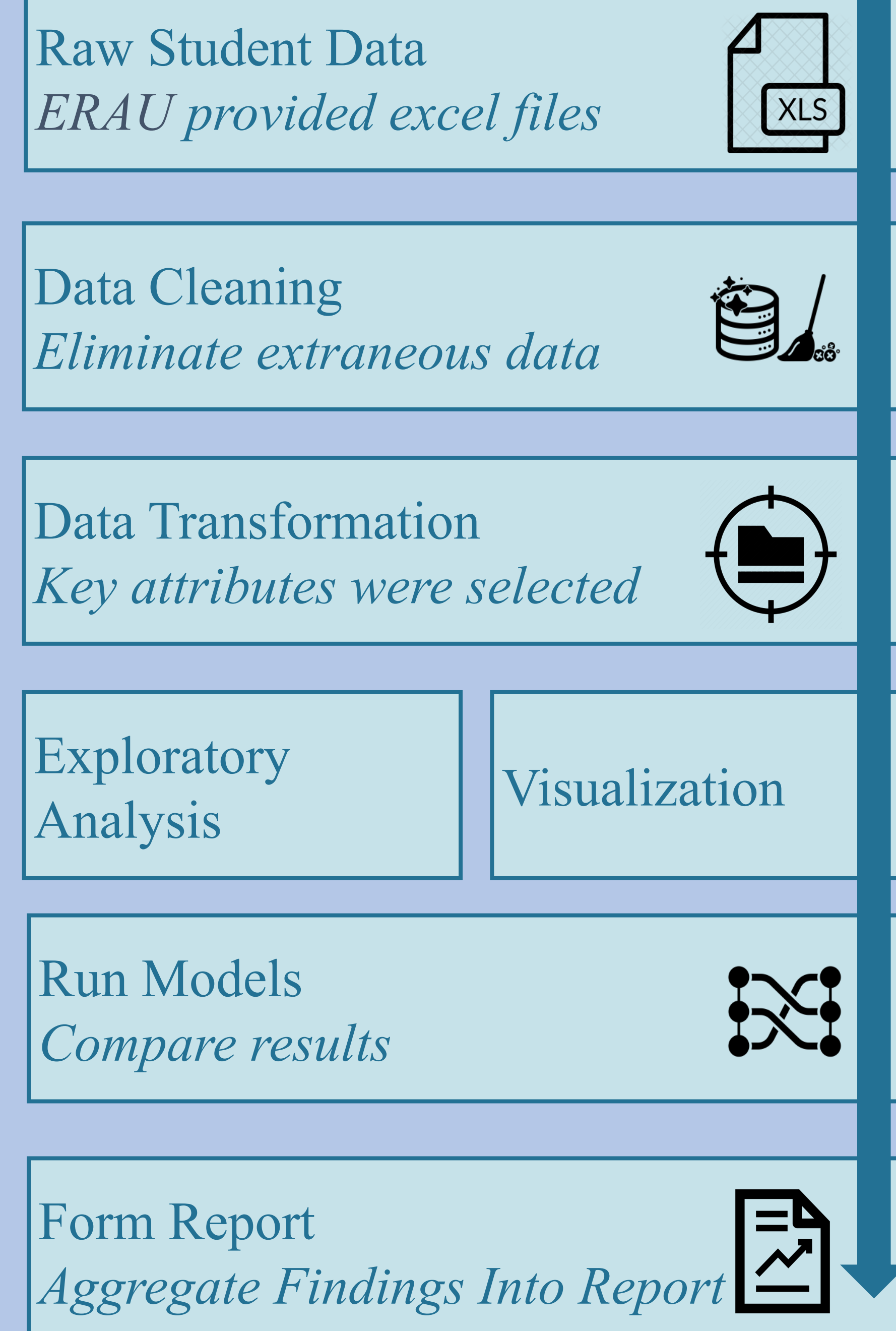
Form Report
Aggregate Findings Into Report

Figure 1: Flow of the research process.

Results

As reflected, ERAU majors and programs are dominated by male student based on gender and white students, which is followed by Hispanic/Latino, and international students based on ethnicity.

	Study Sample	
	n	%
	By Gender	
Male	9453	76
Female	2989	24
	By Race and Ethnicity	
Asian	602	5
African American	555	4
Hispanic/Latino	1817	15
White	7432	60
Two or more races	616	5
Unknown, others, non-resident	338	3
Non-resident	1,087	9

Table 1: Student Demographics

Model

The following four machine learning algorithms were used for predicting major switch in Rapid Miner:

(1)Decision Tree, (2) Random Forest, (3) Gradient Boost and (4) Deep Learning.

ML Algorithm	Accuracy	AUC	Classification Error	Recall
Decision Tree	74.50%	0.50%	25.40%	0
Random Forest	73.70%	0.58%	26.30%	1.10%
Gradient Boost	73.9%	0.57%	26.3%	3.4%
Deep Learning	74.60%	0.64%	25.40%	0.20%

Table 2: Machine learning algorithms results

Among the four ML algorithms, Deep Learning yielded the better result compared to the remaining three algorithms. Factors such as age, admission scores, high school and undergraduate GPA, financial variables were found to have significant correlation with switching majors.

Summary

- The study aimed to reveal key background variables to predict the likelihood of a student switching majors.
- Given the accuracy, area under the curve (AUC), and precision results, deep learning yielded the best result.
- The confusion matrix implies that the algorithm is effective at predicting if a student will switch majors.
- Top correlation factors found for major switching were age, admission scores, and GPA.

Limitations

- Different imputations methods will be used.
- Hyperparameter tunning will be used to improve the model accuracy and fit.