Group Members:

Adrian Alexander, Shaun McKellar, Victoria Zangen

Subject:

IST 652 Spring 2024 Final Project Proposal

Source:

<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention/data>

Topic of Investigation:

Predicting Students’ Dropout and Academic Success Rates

About the dataset:

This dataset provides a glimpse of students enrolled in a foreign university in different undergraduate degrees that are offered at the college. The data set comes in the form of a csv file, and it includes a total of 35 columns in total, along with 4,424 rows of data, providing an in-depth overview of students in the program and their educational outcomes. This overview includes different demographic and socioeconomic factors that could impact a student’s performance at the school. With this critical information, this data can be used to estimate student performance at the end of each semester by assessing the units credited, /enrolled/evaluated/approved and their grade.

On the economic side, to see how external factors could be affecting the students' performance the data set includes critical data to get a better picture of students financially. This is done by including variables such as unemployment rate, inflation rate, and GDP from their respective regions. Analysis of these demographic and socioeconomic factors will provide a deeper understanding of why students are dropping out at such an alarming rate at this university.

Possible Method of Analysis:

* Descriptive statistics by attribute
* Do demographics contribute to a student’s success or retention rate?
* Are there correlations between demographics and social-economics with academic performance? Course studies and academic performance?

Areas needing assistance:

1. Machine learning in python

We intend our project to be larger in scope than the homework by considering predictive factors that may influence or contribute to a student’s retention or dropout rate. Throughout our analysis, it will also be important to understand any biases or risks present in the data.