Predicting the Number of Awards Based on Institutional Characteristics

A Project Proposal Written by Paminas Mayaka

5/23/2024

Problem Statement

What opportunities exist for developing a model to predict the number of awards given for every 100 full-time undergraduate students by colleges and universities in the USA based on various features of the institution?

> ***Target Variable:*** Awards Per 100 Full-Time Undergraduates
> ***Features:*** Institutional characteristics such as 2- or 4-year institution, private or public, student count, number of full-time students, etc.

Context

In the United States, a significant number of students opt to pursue postsecondary education. These students navigate a diverse range of institutional contexts, including considerations such as the types of degrees offered, institutional control (whether private or public), and the availability and quality of financial aid. Choosing a specific institution can be a complex and time-consuming process for both students and their families[1].

Between 2010 and 2021, the total undergraduate enrollment in degree-granting institutions experienced a notable decline of 15%. Interestingly, the largest decrease (8%) occurred prior to the COVID-19 pandemic. This decline in enrollment numbers can have implications for an institution's financial stability[2].

To address these challenges, I propose a model that aims to ease decision-making for prospective undergraduate students and their families, as well as provide valuable insights for policy makers and administrators.

Criteria for Success

- A successful model will predict the number of awards given based on certain institutional characteristics with high accuracy.

Scope of Solution Space

- The focus is on the number of awards given per 100 full time undergraduate students.

Constraints within the solution space

- The data were collected between 2000 -2014. It is possible patterns may have changed especially after Covid-19.

---

[1] Condition of Education 2023
[2] Condition of Education 2023

- The data lacks students characteristics such as gender and ethnicity that may likely influence how awards are distributed.
- Data may be biased.

Stakeholders

- Students: make informed decisions when selecting colleges.
- Policy makers and administrators: make informed decisions about funding initiatives and other education policies as well as assess the effectiveness of existing programs regarding resource allocation and student support.

Source of Data

I will retrieve the dataset for this project from “College Completion Dataset” that is available at https://www.kaggle.com/datasets/thedevastator/boost-student-success-with-college-completion-da/data?select=cc_institution_details.csv . This dataset was originally authored by Jonathan Ortiz (see https://data.world/databeats). The dataset has been available for public use without copyright violations. The dataset contains 63 columns and 3797 rows. The features describe demographic information of US colleges and universities including the name, location, 2- or 4-year, SAT scores, awards, expenses per award, financial aid, public or private, etc. Each row contains information about a specific institution. The authors obtained the data from the National Center for Educational Statistics’ (NCES) Integrated Postsecondary Education System (IPEDS) and the Voluntary System of Accountability’s Student Success and Progress Rate.

Deliverables

- A final project report
- A Github repository for the project
- All code in Jupyter notebook