Predicting Student's Graduation and Dropout Rates

Did you know that a staggering 40% of undergraduates in the US drop out of college before completing their degree? That's a whopping 40 million individuals as of July 2021! Not only that, but dropping out can also have significant financial consequences, with college dropouts earning 35% less on average than those who complete their degree. But why does this happen? Is it due to academic struggles, socioeconomic factors, or something else entirely? Our project seeks to answer these questions by conducting a comprehensive analysis of the key factors that contribute to academic dropout and failure in higher education. With this knowledge, we hope to help reduce the dropout and failure rates in higher education, and pave the way for a brighter, more successful future for students across the country.

SMART QUESTIONS

1. Does a student's average grade in their courses have any bearing on their decision to continue or discontinue their studies in the program?
2. Can we identify trends in the timing of student dropouts within specific semesters, and what are the potential causes for these trends?
3. Are there noticeable differences in student enrollment and achievement across various academic fields or programs, and what are the discernible patterns in these differences?
4. Do factors related to socioeconomic status, like household income, job status, or access to scholarships, play a role in affecting students' academic performance and completion rates?
5. Are there identifiable links between students' demographic attributes, such as age, gender, or nationality, and their propensity to leave the academic program prematurely?

DATASET LINK

The data used in this study was sourced from the UC Irvine Machine Learning Repository. It comprises of 4424 entries and 36 attributes. Every entry represents an individual student.

PROPOSED MODEL

Random Forest Classification

GITHUB REPOSITORY LINK

Link: https://github.com/yashkattimani/Intro_to_Data_Science_in_R_Project_2