Capstone Three - Project Proposal

Context:

It is no surprise to hear that a child's education is one of the most important aspects of their life. It teaches them about the world, makes them smarter, gives them a place to develop academically and socially, and helps prepare them for their career. The big question that many stakeholders have asked - parents, educators, psychologists, students - is how to ensure success?

Is it predetermined by genetics and innate brilliance? Or is it determined more by the amount of effort a student puts in and studies? Perhaps none of that matters - and the economic background and resources afforded to the student's upbringing is what determines their success; afterall, the daily environment and upbringing can have a lasting impact on a student's development. How influential are parents and their involvement, or peers and social relationships?

As one can see, there are a wide variety of factors that can shape a student's experience. Which are the biggest factors that affect a student's success? And can we predict which students will and won't pass?

This will help us to focus on intervention plans - first by identifying which students are most at risk of failing, and second by making a plan for them on what they should focus on to become successful.

Criteria for success:

- Identify the biggest factors that influence a student's success
- Predict which students will pass and which will fail

Scope of solution space:

- Portugal
- Demographics
- Academic habits
- Social and familial influence

Constraints:

- This data only represents students from two different secondary schools in Portugal. There is a large imbalance of school representation.
- The target attribute G3 has a strong correlation with attributes G2 and G1 because G3 is the final year grade and it is more difficult to predict G3 without G2 and G1.

Stakeholders:

- Students
- Parents
- Educators and educational administrators
- Education policy makers
- School psychologists

Data sources:

The data follows students of two Portuguese secondary schools and their attributes such as grades, demographic, social and school related features through school reports and questionnaires. They follow the students' math and Portuguese scores. The data is structured in either binary or 5-level scale classification.

Method of Analysis:

After cleaning and organizing the data, I will produce a few different visualizations to help identify any variables of interest, outliers, or other ways this data may be skewed.

After the initial EDA, I will conduct a PCA to identify which components are the most influential. Lastly, I will run a few different models to identify which can best predict a student's success.

Deliverables:

- GitHub repo
- Slide deck
- Project Report