Capstone Two: Project Idea and Proposal

EdTech students have achieved a variety of successes in their studies and beyond. From developing innovative products and services to winning awards and scholarships, EdTech students have demonstrated their commitment to their education and the field of technology.

Problem statement formation:

The problem we aim to solve is to analyze student performance data on an educational technology platform to identify factors that influence student success and make recommendations for improving student outcomes. Specifically, we will investigate which features of the platform contribute to student success.

Context:

The educational technology platform provides an interactive environment for students to learn and practice coding skills. The platform is used by students of various ages and experience levels to improve their coding proficiency. However, the platform's effectiveness in enhancing student learning outcomes has not been adequately evaluated.

Criteria for success:

The success of this project will be determined by the ability to identify and quantify the factors that significantly impact student performance on the platform. We aim to develop a model that can accurately predict student performance based on the identified features. Additionally, the project should provide actionable insights for the platform developers to improve the learning experience for students.

The project will be considered successful if we can identify significant factors that influence student success, develop a predictive model for student outcomes, and provide recommendations for improving student outcomes. Additionally, the code and documentation should be clear, well-organized, and easily replicable.

Scope of Solution Space:

The scope of the solution space will include exploring the dataset, cleaning and preprocessing the data, performing an exploratory data analysis, identifying significant features using statistical analysis and machine learning techniques, and developing a predictive model. We will also perform data visualizations to better understand the patterns and trends in the data. The final deliverable will be a slide deck and project report summarizing the findings and recommendations.

Constraints:

The project will be limited to analyzing the performance data on the educational technology platform. Data privacy will be a significant constraint in this project, as we will have to ensure that student data is adequately anonymized and protected. The data will be obtained from the educational technology platform's database, which may have some data quality issues.

We will be working with a relatively small dataset of 124,517 rows and 19 columns, with missing values and a mix of data types. We will need to address issues related to data cleaning, preprocessing and feature engineering. We will also need to balance the need for accuracy and interpretability in our predictive model.

Stakeholders:

The stakeholders for this project include the educational technology platform's developers, educators who use the platform, and the students who use the platform to learn to code.

Other stakeholders include the student parents and guardians, and the tutors who provide support on the platform.

Data Sources:

The data for this project will be obtained from the educational technology platform's database. The dataset includes student performance data, such as the number of exercises completed, the level of difficulty, and the time spent on the

platform. Additional data sources may include demographic information about the students and the type of exercises completed on the platform.

The dataset contains information on client and student_usage, performance and demographics. The columns include client_id, user_id, first_trial_appointment_date, first_payment_date, os, tutor, job, task_class, average_score, homework_done, paywall_paid, school_name, desktop_enter, nps_score, add_homework_done, call_date, first_visit_date, region, is_big_city.