

CAPSTONE PROJECT 1: PROJECT PROPOSAL

MINH H. LE

1. What is the problem you want to solve?

The problem that I want to solve is finding out which attributes, including student grades, demographic, social and school related features, have an effect on high school student performance and from those attributes, predicting the student performance (final grade).

2. Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

My clients are educational institutions, and parents. Based on my analysis, schools and parents will know which attributes and reasons affect the student performance. With that knowledge, they can help the students to increase their performance.

3. What data are you using? How will you acquire the data?

The data that I am using was collected by using school reports and questionnaires. The data include two datasets (CSV files), one is performance in mathematics and the other is performance in Portuguese. The data are acquired from the UCI Machine Learning Repository.

URL: <https://archive.ics.uci.edu/ml/datasets/Student+Performance#>

4. Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

First, I will clean up the dataset that is provided. Since there are 32 attributes, I need to figure out which attributes are relevant by regression analysis. After knowing which attribute is relevant, I will use machine learning to predict the final grade (student performance). Since the target variable (the final grade) is continuous, I will use regression supervised learning.

5. What are your deliverables? Typically, this includes code, a paper, or a slide deck.

My deliverable will be in the form of final report and slide deck along with code notebooks