Technical Report: Final Project, Data Science 5110

Felipe Quiroz and Kevin Cooper

December 10, 2024

1 Introduction

A long tradition of friend groups and fraternities alike is tutoring fellow members. Helping study for tests is both a bonding experience and something that helps both people learn and review the material. It is a shame, then, that trying to set up tutoring is often hard. Techniques like "asking around" and "texting the group chat" might work if the right people are paying attention but are not certain. Our project, then, seeks to set up an easy way for people to share which classes they have tutored and what materials they have for those classes. People seeking help can then easily go into which class they need help for and see availability and links to any available resources.

2 Literature Review

The existing body of tutoring services for college students is not ideal. As mentioned above, word of mouth is not a reliable source of discovery, and other services tend to be paid. For example, Knack is a tutoring service that claims to be able to "improve your grades" with its "support network", but what could be more supportive than your friends? It is fundamentally a company with employees, and does tutoring as a job, not as a relationship.

3 Methodology

To complete our project, we broke it down into two parts. One was the data collection and storage, and the other was data visualization and website creation.

3.1 Data Collection

For the data collection, we initially debated many ways of collecting data before settling on Google Forms. Forms has the advantage of being fairly universal, and many people have experience with filling one out. It also automatically outputs its information into a Google Sheet, making the problem of data storage much easier (as the questions can be configured to the database design and vice versa).

3.2 Data Preprocessing

Another advantage of using Google Forms is that it is easy to restrict the allowed input of users to clean values that work with our data. As the code used in our analysis was object-oriented, it is also easy to analyze whether values 'make sense' in their positions – for example, tutors with NULL values for availability are considered not available for tutoring.

3.3 Analysis and Presentation Techniques

To present our data, we use a website hosted on GitHub Pages. It is free and extremely easy to use. The analysis side of our project was in re-creating the data connections from our database in code. For example, when taking in the data from our database, we need to make sure that each tutor is matched with the correct class so students do not try to schedule sessions with people who have never taken the course before! Python is used to automatically scrape data uploaded from the database and create the websites, for which the Dominate library is used. Once the clean data is collected,

Dominate loops over every course and generates a website for that course displaying the tutors, their availabilities, their contact information, and relevant materials. It also creates an index page with a list of all courses that the database has a tutor or material for.

4   Results

The result of all this work is a website that automatically generates a list of tutors from just a Google Form. The index page is a simple list of courses that the database has tutors or resources for, while each course page is a list of tutors (and their contact information and availability), along with a list of hyperlinks to resources. The website is extremely flexible and easily re-generates pages after any given update to its data.

5   Discussion

This website is more able to connect students with tutors than the solutions discussed before. It solves the problems with peer tutoring by being an easy central resource for both tutors and materials, along with showing availabilities centrally on the website. It solves the problems with Knack and other paid services by being free and easy to run with just a Google Form. It also does not require an account or sign-up to use, meaning anyone with an internet connection can just connect to the link and make things work. An issue with our solution is the need for hosting through GitHub Pages, which (reasonably) puts website view limits on its free pages. However, each club would host their own version of this website, meaning these views would be split up and no individual club will hit that cap.

6   Conclusion and Future Work

Some future work we are going to do is connect with more clubs and academic groups on campus to see how this would fit their needs. We will also further develop resources and instructions on how to use this system to best fit the needs of each individual club. A final piece of further work we could do is scaling our system both vertically and horizontally to handle both more data and different types of data.