Intelligent Review System (IRS) Machine Learning V2 Project Proposal

Group Members and Skills

- Jared Raiola
    - Computer Science - Third Year
    - Programming Experience: Python, Java, SQL, JavaScript, HTML, CSS, PHP, C, C++
    - Task: Determine keywords and graph data based on the question classification to help professors and TAs determine which questions to review
        - Create the final review test based off the difficulty scale and keywords
- Adam Chau
    - Computer Science - Third Year
    - Programming Experience: Python, Java, SQL, JavaScript, HTML, CSS, C#, C
    - Task: Utilize more machine learning techniques to help show the data of the questions besides utilizing just score and duration
- Seungju Jason Lee
    - Electrical Engineering - Third Year
    - Programming Experience: Java, C++
    - Task: Determine the difficulty scale based on average time spent, the average grade, number of skips for the question, etc
        - Learn and utilize machine learning to graph the relationships of the questions besides score and duration

Project Goals

- Become sufficient in writing in Python, managing code using Github, and using SQL
- Continue working with Python, SQL, and Ubuntu
- Continue to find and utilize more machine learning (supervised and unsupervised) techniques to help show the data of the questions
- Determine keywords and graph data based on the question classification to help professors and TAs determine which questions to review
- Create a difficulty scale based on the average time spent, the average grade, and the number of skips for the question
- If there is extra time, create a review test based off the difficulty scale for students

IRS-MLV2 Description

The main purpose of ITS is to help students succeed in the class, and all of the resources and data provided in the system are targeted towards benefitting the students themselves. The team believes that TAs serve an integral part in a student's success and we believe they need additional insight into how the class is performing. Last semester, the IRS-ML team used clustering techniques such as k-means clustering and agglomerative clustering to show the relationship between score and duration. Our goal is to develop a system that will help TAs identify the subject areas in which the students are struggling in order to provide a more personalized learning experience to each recitation section of Digital Signal Processing (DSP). Interesting ideas to research would be unsupervised machine learning techniques and see how the technique differs from the clustering. A potential pitfall would be learning exactly how all these machine learning techniques work.

IRS was created as a data visualization tool for TAs to help them review what questions and concepts proved challenging to the students. This semester, the team plans to use more machine learning techniques to help understand relationships besides score and duration for each semester such as sorting the data by keywords. The primary goal of the semester is to determine keywords and graph data based on the question classification to help professors and TAs determine which questions to review. Another goal is to determine the difficulty scale based on average time spent, the average grade, number of skips for the question, etc. This will give TA's a more accurate image on what to review with students. Currently, the last assignment on the ITS is a combination of all topics covered throughout the semester. The final goal of this project, if time permits, is to take all of this data and formulate a review test, personalized to each student, containing the topics and questions they struggled with.

Project Timeline

Week 1 - 4	- Split Into 2 Teams - Make a Project Proposal Idea - Create Project Proposal - Successfully download Ubuntu - Project Proposal Draft Due
Week 5	- Final Project Proposal Due - Start to learn SQL, Python, and Machine Learning - Set up SQL- become acquainted with SQL/Python - Look into scikit-learn and other libraries for ML

Week 6 - 8	● Determine whether to use supervised or unsupervised learning ● Create an algorithm based on keywords for questions, which determines overlapping/related questions
Week 9 - 12	● Determine question difficulty based off the algorithm we made for the keywords and students' relative performance ● Create a difficulty scale based on the average time spent, the average grade, and the number of skips for the question
Week 13 - 14	● Develop an algorithm to create a review test for students based on question difficulty if time permits ● Start to upload the final code on GitHub ● Start making code explaining the code to future VIP members
Week 15	● Finish making comments on code explaining the graphs and project for future VIP members ● Upload the remaining code to GitHub ● Work and prepare for the final presentation
Week 16	● Present the final presentation

Implementation Tools & Resources
- GitHub: https://github.gatech.edu/VIP-ITS
- Ubuntu
- SQL Library
- MySQL Workbench
- Project Documentation Notebook
- Machine Learning Library
- Python Library