CS 410 Project Proposal
Recommender System for Music/Songs

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   I will be completing this project by myself. My netid is shengya4 and my name is Arthur Huang.

2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

   My free topic is building a recommender system for music/songs. In general, recommender systems are seen as a "push" access mode in text information access. In other words, it recommends content to users.

   This project will explore content-based filtering. It is interesting because recommender systems are becoming more prevalent in modern platforms. They are utilized in apps such as Netflix, Instagram, Tiktok, etc.

   My planned approach is to first explore some datasets then process them. Next I will implement several critical components: **a classifier (binary) which decides if a movie would be recommended to a user, and a function to initialize some user profile, if time permits, a function to respond to relevance judgements.** Some helpful NLP libraries include spaCy and NLTK. For data processing and computations for machine learning methods, Pandas and Numpy, scikit-learn will be helpful.

   I'm still exploring some datasets but an example one I likely would use can be seen here: https://www.kaggle.com/code/vatsalmavani/music-recommendation-system-using-spotify-dataset

   The final product should be a system which can be run (likely in the form of an executable script). Given some user information, some relevant songs/music would be recommended.

   Which programming language do you plan to use?

   Python will be used for this project.

3. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

   Explore, clean dataset -> 5 hours
   Implement initialization of initial user data -> 5 hours
   Develop functions for comparing content similarity -> 5-8 hours
   Develop the binary classifier -> 5-8 hours