

1. Which tasks have been completed?

The first milestone of the project is to find a suitable dataset/suitable datasets. This part has mostly been completed. I searched the UCI machine learning repository for datasets not only limited to recommender applications. Some appealing ones I found were related to lyrics, texts, and songs in general. I have also successfully loaded these datasets using libraries such as Pandas and Numpy, and did some initial exploratory data analysis.

I have also completed parts of the next milestone, which is defining a binary classifier. I did not jump straight into implementing such a function for my specific datasets, since I believe it is good to develop a good understanding of similar algorithms first. I've completed scratch work with logistic regression classifiers on experiment data.

2. Which tasks are pending?

For the next few weeks, the tasks that are pending are refining datasets. Although several have been picked, I would like to get a sense of which one would work best with my project. Additionally, I need to complete functions to compare content similarity for songs and functions to classify songs to either "should recommend" or "should not recommend". After reviewing the proposal feedback, I realized I needed to also add some novel ideas to the project, on top of it being a simple recommender. I am thinking of allowing users to define certain sentiments and picking out specific lyrics.

3. Are you facing any challenges?

The main challenge I am facing is scoping out a novel idea to add to my project. Currently I am brainstorming and researching several different ideas. However, I believe the

most feasible and interesting one would be recommending lyrics based on user defined sentiments or other factors.