

CS 410 Final Project Progress Report

Predicting Song Genre Based On Lyrics

1. Team Members

Aaron Aftab (aaronaa2) - Team Captain

Avinash Nathan (anathan4)

2. Progress Made

By leveraging the Spotify API (<https://developer.spotify.com/documentation/>), we have been able to pull data regarding songs and their genres. Since Spotify songs are not specifically associated with genres, we used the first-listed genre associated with the corresponding artist. In order to get lyrics data, we used the Genius API (<https://docs.genius.com/>) to access relevant lyrics pages and scrape the desired text. Now, we have a full dataset containing songs, their associated genre, and their lyrics.

3. Remaining Tasks

To complete this project, we will need to create the pipeline for modifying and processing our lyrics data (stopwords, stemming, etc.) and then test different models with our separate lyrics data set to find which is the most accurate. Then, we will build the frontend allowing a user to input a set of lyrics and receive a genre categorization.

4. Issues Faced

We have not encountered any major issues yet - data collection via the Spotify API has been smooth and we have been able to successfully generate a varied dataset, as we desired. We initially ran into an issue when we realized that the Spotify API did not provide lyrics, but we were able to find another API (Genius) to allow us to do this.