

Personalization and Machine Learning  
Assignment 1 - Report  
By Marissa Beaty

For this Assignment, I analyzed one of my own playlists called “An Anthology of My Favorite Songs.” I selected this playlist as it contains only songs I know and love, which provides me with the knowledge to accurately determine the model’s recommendations and reliability.

When working through this assignment, I was most interested in how the songs on this playlist relate to one another, such as what features these songs might share? What songs are more related than others? Most importantly, how well can the model recommend songs to me? Given these interests, I will discuss Task 3 and Task 4.

For Task 3, I selected a song that doesn’t quite fit into the rest of the playlist: Billie Holiday’s “All of Me.” Though the playlist contains a large variety of artists and styles, “All of Me” is the only explicitly deemed jazz song and the only song from a different generation than my own. To begin the task, I started with a setting of  $N = 10$ . To get a better idea of the results, I created a features table of the 10 recommended songs and the initial song to see the technical similarities. My initial reaction to the model’s recommendations was quite surprising, as I wouldn’t have recommended these songs myself, however, after reviewing the features of each song, I began to understand. Though vastly different in genres, each song was similar in features. In fact, all recommended songs have an energy no greater than 0.5140 and, despite the outlier of Kevin Garrett’s “How Dare We Fall,” have an acousticness higher than 0.673. It should be noted that, where Kevin Garrett’s differs in acousticness, it is nearly identical to our initial song in danceability and shares quite close resemblances in liveness and valence. What appears to be the trend in these recommendations is a wholistic review of all features and that relationship with the originally selected song.

When we increase the value of  $N$  to 20, however, my confidence in the recommendation system falters. There remains a focus on the instrument, but the sentiment of the song no longer fits. This is exacerbated even further when we increase the value of  $N$  to 40. With a dataset of only 50 songs, we cannot blame the model for making inaccurate recommendations. Should our sample have been larger, say 500 songs instead of 50, I would expect a recommendation of 40 songs to remain quite consistent with the accuracy of our recommended 10 songs. As that is not the case, I would recommend basing the  $N$  value on 5-20 percent of the overall size of the dataset.

To investigate the other question of Task 3, I removed the features that seemed most related in our previous tests. That is: acousticness, energy, liveness, valence. The results of this are quite interesting. For one, several songs still appear on this new list as the list before, including but not limited to Taylor Swift’s “Sweet Nothing” and Mt. Joy’s “Bathroom Light.” Where “Sweet Nothing” was our first recommended song prior to changing the features, it was replaced by The 502s “Just a Little While.” This confirms that some songs were recommended for the removed features as suspected, but not all. This reinforces the hypothesis that there is a holistic approach to the recommendation system.

The other task I decided to analyze is Task 4. What is interesting about this is referring to my selected song in Task 3, Billie Holiday’s “All of Me.” In this view, it confirms that the song stands alone, but so do a few others, most notably Catfish and the Bottlemen’s “Tyrants.” The rest of the songs on this list, however, are clumped together quite nicely. I would not necessarily

consider this a good way of recommending songs. For example, in space Anderson .Paak's "Jet Black" is very close to The Lumineers "Where the Skies are Blue." Thematically, these songs are vastly different and when I am in the mood to listen to one, I am not in the mood to listen to the other. If anything, this perspective reinforces how diverse this playlist is, and how this diversity might confuse the algorithm, or how the algorithm categorizes songs differently than I do. Despite this, when keeping the N values at a correct proportional value to the size of the dataset yields quite a successful model, certainly not perfect, but limitedly successful, nonetheless.