

Noah Falanga

Self Assessment 2

Part A:

My individual contribution to the assignment was originally supposed to be the construction of the machine learning algorithm to be implemented in our final product to cluster users, and I ended up fulfilling this goal, albeit to a different degree. Originally, our scoring functions between users were to be constructed through a variety of features, such as common songs between each other, individual song attributes provided via Spotify, artist of the songs, and more. Spotify, however, deprecated some features of its API, introducing an obstacle in receiving crucial song attributes that we intended to use in our algorithm. Because of this, we had to pivot to using genres to build our user vectors, with each feature on a user vector correlating to a specific supergenre. After working with the available data about the songs we created, I was able to produce and explain a clustering algorithm that fits users together based off of supergenre distribution, with each feature being the proportion of songs belonging to a given supergenre in their current 100 most listened to tracks.

In terms of code, I created a file that tried out and recorded the results of multiple hyperparameter configurations of a few clustering algorithms on a variety of forms in which our preprocessed data took. I then took the best fitting clustering algorithm and implemented it into the backend of our application. I trained the classifier on initialization, assured all users in the database had the correct label according to the classifier, assigned labels to new users and put them in our database upon login, calculated individual user's distances from other cluster centers, and finally used those distances to select from which clusters should songs be pulled upon for playlists that are made from inside the user's cluster, made from the closest few clusters, and furthest few clusters. I grew in the identified skills from the initial assessment. Using machine learning libraries with a dataset that was collected independently rather than acquired and already preprocessed from a third party was a valuable lesson in data science. I also was able to use a clustering method in a way that was not just descriptive, but also generative. While I did not build my front-end experience, I did grow in web development skills, namely with balancing database requirements and pings such that data was consistent without compromising performance. While I did just utilize simple clustering and did not rely on more complicated, nuanced artificial intelligence methods, I still deepened my understanding in the field through the training and interpretation of the clustering algorithm.