**Data-X Final Project: Spotify Data**

Ellek Linton, Daron Assadourian, Holly Cheek, Tianchu Hang, Alessandro Buy, Alan Zhang

This semester, our project aimed to improve the quality of music listening. At first, we wanted to build a machine learning model to actually quantify song features (like happiness, energy, etc) from an audio file. We realized Spotify already offered this through their API, so we instead focused our efforts on improving music recommendation through utilizing the Spotify-computed song features. Ultimately, we created an alternative music recommendation service that fixes many of the problems in current collaborative-filtering based recommendation systems.

We created a music recommendation program that recommends better music that is currently offered by the leading collaborative filtering algorithms. The first step in this process was to compile a large database of songs on Spotify, complete with Spotify’s quantitative song features like *happiness* or *acousticness*. After doing this, we cleaned our data and created a searchable tree structure to find “similar” songs given an input song (we used *SKLearn.BallTree*). After doing this, we learned how to integrate our program with the Spotify API to allow real-time search results using a query as inputted by the user. After getting the input song from the user, we search our *BallTree* for the user-specified *k*-nearest tracks. From the user’s perspective, they must first authenticate their login with Spotify, and then search for a “seed” track. After configuring playlist parameters like year-range and number of tracks, a playlist of similar songs is presented to the user.

This solved several problems related to current music-recommendation algorithms. Most importantly, the current collaborative-filtering based recommendation algorithms prevent users from truly discovering “new” songs. In researching this space, we discovered that many users were unhappy with the repetitive nature of music recommendation within Spotify--after about 50 tracks, you hear the same songs over and over again. Since our algorithm uses a quantitative nearest neighbor search of over 400,000 tracks on Spotify, the recommendations obtained through our program are much more diverse--they are different for every input song! This allows users to truly discover new songs that they would have never otherwise have heard of.

In our program, we also incorporated a tool to visualize the user’s music history: the songs that the authenticated user has “liked” on Spotify. By integrating with the Spotify API, we are able to visualize *happiness*, *acousticness*, and several other quantitative features and how they have changed over time. This allows users to discover more about their own personalized listening trends and gain deeper insights into the seasonality of the music that they listen to. In our research, we also discovered that users felt they did not have clear access to their Spotify data; this solves that problem. Users felt like they did not have access to their Spotify data and may not necessarily understand their music. By using our program, users can obtain a comprehensive understanding of how their musical taste has changed over time and receive more powerful music recommendations.

**Resources**:

Final Presentation Slides:

<https://docs.google.com/presentation/d/1JMQqxOU-P4ZWEMg5_X-ODZvkbuqGSocFE9-N6D3mAp0/edit?usp=sharing>

Link To Shared Google Drive:

<https://drive.google.com/drive/folders/1y0_1Cy2ZLzMTq_LkEsbOB4M-E5nmD6Sg?usp=sharing>