Creating the Perfect Playlist: Content-Based Generation of Spotify Playlists
Sonu Mehta, Laura Ware, Omar Abboud

GOAL
Explore methods for predicting the success of a Spotify-curated playlist, based only on data about the songs that comprise that playlist, and use these models to develop novel processes for curating successful Spotify playlists.

Features:
- Predict the success (# of followers) of a given playlist
- Identify songs that are acoustically similar to a seed song
- Generate a playlist that optimizes the grouping and sequence of tracks

DATASETS
Only a small number of playlists have over 100k followers

Spotify API
Includes data about each Spotify-curated playlist (e.g. total tracks, sequence, no. of followers) and individual tracks (e.g. audio features, popularity)

Raw Audio
30-second samples available from the Spotify API for ~25% of the 20k tracks in our dataset.

MODELS

Random Forest to Predict Number of Playlist Followers
Inputs: Acoustic features (danceability, loudness, energy, liveness, etc.), duration, popularity of songs, track order
Output: No. of playlist followers, divided into 5 bins

Using Raw Audio to Predict Track Popularity
Inputs: MFCCs, Chroma coefficients, Energy at each timestamp
Output: Popularity divided into 6 bins
Performance: Random forest classifier gave 75% accuracy, F-1 score of 1 for the top class

CONCLUSIONS
- Combinations of songs can be reasonably optimized for popularity using simulated annealing supported by raw audio-based similarity metrics and a Random Forest predictive model of popularity.

Simulated Annealing increases optimization performance significantly

RESULTS
Classification Model Comparison
The mean popularity of a playlist's tracks alone is not sufficient to predict overall playlist followers; this metric is much stronger when combined with acoustic features.

<table>
<thead>
<tr>
<th>Predictor Set</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Popularity Only</td>
<td>0.35</td>
</tr>
<tr>
<td>Spotify Acoustic Features Only</td>
<td>0.66</td>
</tr>
<tr>
<td>Mean Popularity + Acoustic Features</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Weighting of Top Tracks vs. Performance:
Using a higher percentage of top tracks improves performance but decreases the acoustic similarity of the resulting playlist.

REFERENCES