

Content-based Prediction Of Music Taste

Juan Pablo Brasdefer

Fabian Pawelczyk

2023-05-22

Contents

1	Abstract	1
2	Proposed Method	1
2.1	Experiments	1
2.2	Acknowledgments	1

1 Abstract

Results between Logistic, XGBoost, and RandomForest models are assessed, with the RandomForest ($F1 = .67$) proving strongest.

2 Proposed Method

2.1 Experiments

Data: This project utilized song data taken from directly from Spotify playlists. Every year, Spotify releases a *top-100 Songs* playlist for each of their customers that provides their most listened songs of that year. Playlists ranging from 2016 to 2021 were used for the three individuals in the group, and the playlists were chosen because they are the most efficient way of illustrating the music tastes of each individual. In order to pull this song data into Python, Spotipy was used. Spotipy is a lightweight Python library for the Spotify Web API. With Spotipy, you are able to gain full access to all of the music data provided by the Spotify platform.

Experimental details: We decided to proceed with three different classification models to work with our newly created Spotify data that contained information on the three individuals. These were: Logistic Regression, XGBoost, and Random Forest. Each will be discussed in details in the following paragraphs.

Results: As mentioned during the Experiment Details

2.2 Acknowledgments

One of our initial inspirations for this particular project came from Adam Reevesman's article on Towards Data Science¹. It provided a framework for how Spotipy can be used to gather information about an individual's song preferences and encouraged us to pursue this idea further. Additionally, the code he utilized to pull and clean the data from Spotify was essential in ensuring that we had a well structured data file to work with as we built out our three models.

¹<https://towardsdatascience.com/spotify-wrapped-data-visualization-and-machine-learning-on-your-top-songs-1d3f837a9b27>