

Capstone 3 Report

Nathan Knudsen - Nutrition and Weight Loss

Explain The Problem

Music is a massive industry, with artists all over the world constantly vying to know how to best cater to their fans, build their brand, and remain relevant. In this project, we explored data pertaining to a particular music producer and DJ by the artist name of Illenium. In exploring this data, we are hoping to maximize Illeniums popularity (and subsequently profits) by analyzing the feature selection of his songs.

My Approach

Using the API for Spotify, we pulled in the 20 artists most similar to Illenium as well as their top 10 most popular tracks. By doing this we are sticking with similar genres and music choices to cater to the target audience for this artist. Each track had features related to it, such as danceability, valence (songs happiness), tempo etc. From these features we built supervised machine learning models to see how all of these features would affect track popularity and subsequently the popularity of the artist as well. Once the training and testing sets were built and our model was selected, a new dataframe was created with randomly generated feature values as well as the predicted popularity based off of the features to see if we could rapidly simulate the track creation process and maximize popularity rapidly.

My Findings

Our selected model was a Random Forest Regressor model, which has an accuracy score of 86%. The randomly generated dataframe maxed out popularity at about 58 (which was significantly lower than the original datasets max popularity of 76, and was a lot closer to the mean and median of 49)

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature	pred_popularity
204	0.39	0.27	5.46	-3.24	0.32	0.34	0.34	0.02	0.68	0.78	83.63	1.73	58.53
107	0.44	0.07	1.22	-5.97	0.40	0.34	0.08	0.26	0.35	0.67	123.26	2.74	58.18
617	0.35	0.14	7.95	-14.29	0.69	0.33	0.44	0.00	0.62	0.88	115.90	3.29	58.13
397	0.83	0.16	2.21	-9.13	0.50	0.33	0.00	0.06	0.64	0.49	89.43	2.51	58.13
500	0.40	0.26	4.00	-2.41	0.71	0.21	0.14	0.32	0.69	0.50	143.99	2.80	58.10

This is interesting and makes sense as well, since the popularity of a song is incredibly subjective and highly influenced by variables that are hard to quantify such as how recent the song is.

Ideas for Further Research

I would be interested to see how the model would react if even more songs were added (by either expanding the artist pool or including all songs from included artists). The API for Spotify was incredibly vast as well, so there could potentially be so many other applications or features that could be experimented with.

3 Recommendations

1 - Try using the feature values from the first row of the prediction dataframe - If Illenium wanted to experiment with creating some songs with these feature values, the model would suggest that he could create tracks more popular than average (and could save some of the song creation time). The model suggests that he would get a popularity score of at least 58

2 - Collaborate with other artists high on the popularity list - I found through the analysis process that many of the most popular tracks were collaboration efforts between popular artists. This increases the chances of having a popular song!

3 - Ignore the model and artistically create music - By simply creating music and doing what he loves, Illenium was able to create a track with a popularity score of 76 (without the use of analysis or models). Music and Art is such a subjective topic, that it would suggest that artists on the higher end of the popularity spectrum have already figured something out and can continue to artistically create content. This project and model might be more useful to artists who are on the lower end of the popularity spectrum and are struggling to get by, as it could give them a roadmap on things they could do to increase popularity.