Springboard

Capstone 2 Proposal for Ashley Owings

What is the business problem?

When marketing music, it is imperative to classify the music into a genre that is both popular (so as to have a larger potential audience) and well matched so that the market/audience is likely to connect to the music. By having a good classification methodology, it makes it more likely to find the appropriate audience. It also allows previously recorded music to be classified to current musical tastes and genre definitions as they may have changed since the music was first released. The focus of this project is to identify the key factors that classify the genre of a song and create an algorithm that will define the genre based on those factors.

Who are the intended stakeholders, and why is this problem relevant to them?

Intended stakeholders are musicians, A&R, and other associated music professionals. It will allow them to place the music in the most appropriate market to allow the music the most opportunity to be successful in the market.

Where are the datasets available from?

In this project, we will use a Spotify dataset already pulled from Kaggle that will allow us to search the data for genre classifications.

What data science approaches do you anticipate you will use to model the business problem as a data science problem?

To predict that a song is a specific genre requires machine learning classification models, a number of features defined by the Spotify API will be added to the data set to determine which features correlate to the classification of genre. Potential features include: danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration, and time_signature.

We will use the supervised learning decision tree models which can also determine feature importance automatically.

How do you anticipate that you will evaluate the performance of each of the data science approaches that you currently envision?

To identify the best model, we will look at specific statistics to indicate the fit of the model. We will want an accuracy, precision, recall, and F-1 statistic to be as close to 1 as possible to indicate how often we get a correct prediction from the model.

Certain features may be more important to the model than others. We may want to include some of these statistics when looking at their impact to the result when all other variables remain equal.

How do you anticipate that the intended clients will use the results of your CP2 to address the original business problem?

When listening to new music by current or upcoming artists, the algorithm can be used to make sure the songs are well matched to their genre - and each other. If multiple songs by an artist are wildly different in genre, it might suggest that they are not well suited to be included together on the same album. Suggested genres resulting from the album will identify which markets the songs and albums will be placed and marketed into.

Deliverables

- Jupyter notebooks
- Final report
- Presentation slide deck