Genre Genius : Predicting Genre Given Audio Features

INTRODUCTION

When you're listening to music it's easy to notice your foot bouncing to the rhythms, your head bobbing to the beat; later in the day you might find yourself humming a tune over and over again. What is it in songs that please our ears so much? Are the elements of songs we find so intriguing common among all songs, across all genres? That is, are all songs fundamentally the same? By first overcoming a variety of obstacles, then centering our focus on two, ostensibly distinct genres, we explored this question by creating a kNN prediction model.

The goal of this project, then, was to create a model in which the audio features of a song (ex. decibels, tempo, danceability, key) could be input and the genre of the song could be predicted. In this way, songs with unspecified genres in Spotify, for example songs produced by smaller artists, can be inferred by our model. Having a specified genre might help in getting the smaller artist more exposure, as they could be recommended to listeners who prefer that genre.

BACKGROUND

Initially we hoped to produce a model in which song popularity could be predicted given audio features. We decided to abandon this idea because we found weak correlation between various audio features and popularity. The main obstacle here was the fact that over time the types of songs that have been popular have changed drastically – for example, in the 70s, rock was super prevalent **[1]**, whereas in the 2000s, pop took center stage **[2]**.

Thus, we switched gears to focus on predicting song genre given audio features. With our new focus in mind – predicting genre rather than popularity – we aimed to consider and classify as many feasible genres as possible. This proved to be too overwhelming to the model. We found no distinct correlations between genre and audio features. Next, we considered a dataset with less overwhelming data – less genres to consider. Once again, we found no unique correlations between genres and audio features. Our kNN model had a prediction accuracy of 50%. This led us to conclude that perhaps all songs, on a fundamental level, are the same, particularly of the Dark Trap, Emo, Hiphop, Pop, Rap, R&B, Trap Metal, and Underground Rap genre types.

With this in mind, we turned our attention to a third dataset, one in which we could compare two seemingly different genres: jazz and rock. The question we hoped to answer with our new model was whether our assertion of all songs being fundamentally the same is true. If this is true, our model would remain at a consistently low accuracy rate. If it were untrue, the accuracy rate should skyrocket.

RELATED WORK

Previous work has been done to predict songs given audio features. In "Using Machine Learning to Predict Song Genre from Spotify"**[3]** and "Predicting music genres using waveform features,"**[4]**, both used a kNN prediction model, resulting in an accuracy rate of a measly 53% and 65%, respectively. Both projects came to the conclusion that they were considering too many genres at once, thus overwhelming the model. Seeing as our project faced the same issue, resulting in a disastrously low accuracy rate of 50%, we turned our attention to other previous work.

One project in particular inspired us to narrow down the focus and create a model that was meant to distinguish between two different genres: "Using K-nearest neighbours to predict the genre of Spotify tracks"**[5].** This project initially approached the model as we did, but reached a similar block of low accuracy. Thus, the project turned its focus to distinguishing between two objectively different genres: jazz and hiphop. This inspired us to focus on jazz and rock ourselves. Rock was chosen over other genres like electronic music because it showed the greatest difference between its audio features.

THE DATASET

All three datasets were sourced from Kaggle, each of which pulled from Spotify's public song data. The first dataset provided 600 top songs from 2010 - 2019 with 50 unique genres; the second provided around 18,000 songs with 8 genres; the third dataset provided around 9,200 songs with 9 genres. As previously mentioned, though, only two genres of this last dataset were considered: jazz and rock. Thus, the primary dataset we manipulated had a total of 2,111 songs. All datasets included 15 audio features, most of which were ranked from a scale of 0-1: energy, danceability, liveness (presence of audience), valence (positivity), acousticness, speechiness (lyrics), popularity, mode (major & minor), and instrumentalness (sounds like "Ooh" & "Ah"); a scale of -60 to 0: decibels (dB); duration in ms; bpm (beats per minute); key (C, C#, D, D#, etc. transformed as 1, 2, 3, 4, 5, etc.).

Cleaning the Data

Upon choosing our primary dataset, in order to minimize bias towards one song and subsequently one genre, we dropped all duplicate song identification numbers. We then dropped all NaN values and all columns we deemed irrelevant to our model: uri, track_href, analysis_url, and Unnamed: 0 (non-audio features). Next, we pulled all rows from the genre column that were labeled as jazz or rock. Finally, in order to be able to manipulate the data into a kNN model, we converted these genres to have a numerical value – 0 for jazz, 1 for rock.

The Variables

In order to distinguish which features were useful for our prediction model, we first created a distribution of mean audio features with each genre using a stacked bar graph (*see Figure 1*). Here, we saw a great prevalence of acousticness in jazz over rock, for example. For audio features like duration, though, we saw a pretty equal distribution between genres.

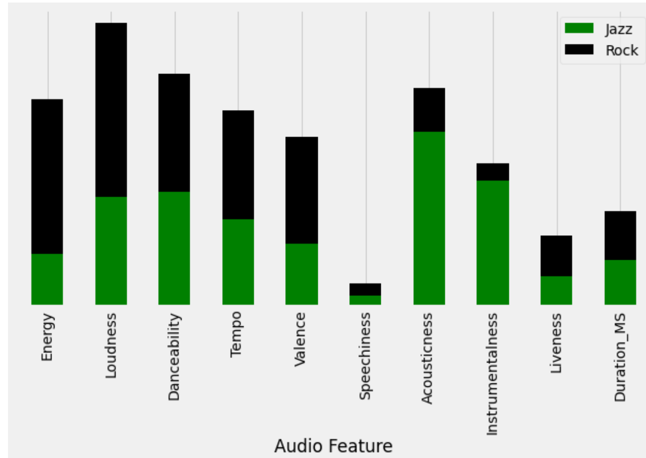

Figure 1: Stacked Bar Chart of Audio Feature Values By Genre

To investigate these relationships further, we plotted histograms considering each variable between jazz and rock. Histograms like *Figure 2*, where there was minimal overlap was determined as useful, whereas histograms like *Figure 3,* where there was significant overlap, were considered not useful. In this way, we concluded energy, loudness, acousticness, instrumentalness, valence, and tempo to be adequate predictors of genre type.

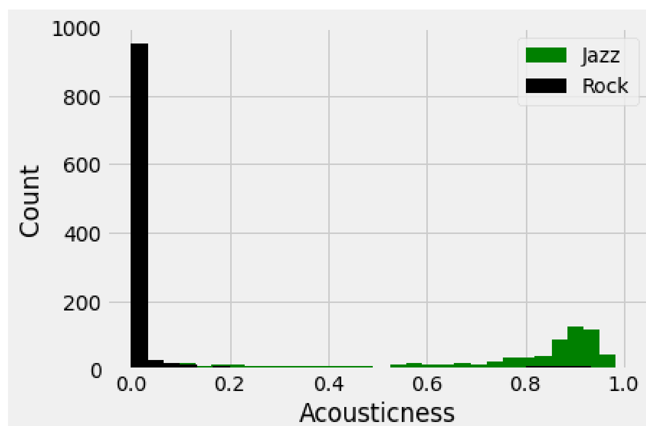

Figure 2: Distribution of Acousticness Between Genres

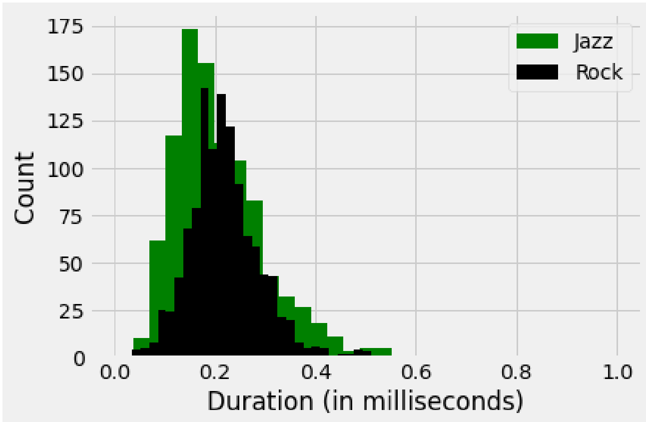

Figure 3: Distribution of Duration Between Genres

THE PREDICTION MODEL

With useful audio features identified, a new dataframe was created with all units standardized. This helped mitigate errors and bias due to different metrics of the audio features. The new dataframe allowed us to accurately compare audio features and the relationships that exist among them. An 80/20 split between training and testing data was created for the kNN model in order to predict if a song belonged to the jazz or rock genre. After creating an elbow plot (*see Figure 4*), we found the optimal k value for kNN to be 5. When k=5, our accuracy rate was 92%.

Figure 4: Elbow Plot for Optimal K value

In order to visualize the decision boundaries of our kNN predictor, we used the principal component analysis (PCA) method (*see Figure 5*). This way we could include all the significant features we identified to show a distinction between jazz and rock. The PCA method was used because we needed to reduce the dimensionality of our data since we have 6 audio features. This was essential because it can keep as much of the original data as possible, while combining them using the covariance matrix and its eigenvectors, allowing us to plot on a 2-dimensional graph. On the x-axis, principal component 1 includes the most significant values and stores maximum information, while principal component 2 on the y-axis is the second most significant with the rest of the information. After researching the issue of dimensionality, we found the PCA method to be optimal for our visualization **[6]**.

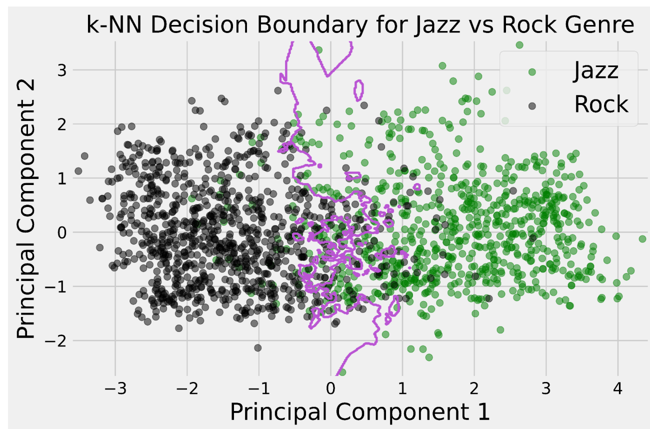

Figure 5: Decision Boundary created by All Significant Audio Features of Jazz and Rock

CONCLUSION

By narrowing our focus to just two, very different genres, we found that our original assertion of all songs being fundamentally the same across all genres was wrong. Considering audio features like energy, loudness, acousticness, instrumentalness, valence, and tempo, our kNN was able to predict genre type with 92% accuracy. Thus, we were able to reveal a clear distinction between at least two genres. That said, we also have to consider that this model only had a high accuracy rate because it was distinguishing between two inherently different genres. It may not apply to all combinations of genres. There are still some instances where there are overwhelming similarities between genres, for example between rock, rap, pop, and trap. Distinguishing between these genres would likely require more advanced models than what we are currently familiar with.

Further, we also have to consider that music cannot be judged solely on audio features. Beyond musical techniques, genre has more factors associated with it, for example cultural context, spirit, and content of the song like the lyrical message **[7]**.

Finally, another aspect of this project we need to consider is the way in which Spotify quantified their data. The way in which these audio features are measured is unclear, especially for subjective values like danceability, energy, and liveness. This may be another reason why it was difficult to find stronger correlations in the beginning stages.

Going Beyond

In earlier discussions we mentioned there was an overwhelming similarity between songs when considering too many genres at once. Even when given useful predictor variables, our kNN model was unable to accurately distinguish genres. Thus, we came to the conclusion that genres like rock, pop, hiphop, and R&B in our first dataset are very similar in composition Looking at our second dataset, we found a similar relationship between the genres Dark Trap, Emo, Hiphop, Pop, Rap, R&B, Trap Metal, and Underground Rap. In this way, since hiphop, rap, and R&B proved to be similar to rock, we can conclude that genres ranging from Dark Trap to Underground Rap are fundamentally similar to rock. In terms of going beyond, something interesting to consider is the ways in which our model could classify these indistinguishable genres. In theory, since on an audio feature level they all seem to be similar to rock, they should be classified as rock rather than jazz in our prediction model.

GROUP CONTRIBUTIONS

While each member contributed to all facets of the project, main roles were as follows: Jared Maksoud was in charge of cleaning the data and creating visualizations. Katja Edwards was in charge of literature review and identifying useful predictor variables. Sabrina Matsui was in charge of the kNN model and mapping the decision boundary.

REFERENCES

[1] Hafeez, A., & FutureSchool, B. Y. J. U. S. (2022, June 22). *Rock music: What is the Best Decade of rock music ?* BYJU'S Future School Blog. Retrieved February 24, 2023, from https://www.byjusfutureschool.com/blog/what-is-the-best-decade-of-rock-music/#:~:text=The%20Best%20Era%20of%20Rock%20Music%20%E2%80%93%201970s,-The%20tragic%20beginnings&text=The%20music%20created%20by%20the,people%20to%20this%20very%20day

[2] Myers, K. (2016, July 18). *The history of pop music in 5 defining decades*. Culture Trip. Retrieved February 24, 2023, from https://theculturetrip.com/north-america/usa/california/articles/the-history-of-pop-music-in-5-defining-decades/

[3] Laso, J. (2020, December 28). *Using Machine Learning to Predict Song Genres from Spotify*. RPubs. Retrieved February 22, 2023, from https://rpubs.com/jsl0516/spotify_genres

[4] Venturott, P. H. G. (2021, March 3). *Predicting music genres using waveform features*. Medium. Retrieved February 22, 2023, from https://towardsdatascience.com/predicting-music-genres-using-waveform-features-5080e788eb64

[5] Thomas, N. (2020, May 9). *Using K-nearest neighbours to predict the genre of Spotify tracks*. Medium. Retrieved February 22, 2023, from https://towardsdatascience.com/using-k-nearest-neighbours-to-predict-the-genre-of-spotify-tracks-796bbbad619f

[6]YouTube. (2019). *Principal Component Analysis in Python | Basics of Principal Component Analysis Explained | Edureka*. *YouTube*. Retrieved February 24, 2023, from https://www.youtube.com/watch?v=n7npKX5zIWI.

[7]Wikimedia Foundation. (2023, February 23). *Music genre*. Wikipedia. Retrieved February 24, 2023, from https://en.wikipedia.org/wiki/Music_genre#:~:text=A%20music%20genre%20or%20subgenre,a%20wide%20variety%20of%20subgenres.