Identifying trends/patterns that influence song popularity on Spotify

Joshua Lin, Jolene Okamoto, Son Nguyen, Ji Yoon Rhee

Northeastern University

ABSTRACT

This project will focus on identifying trends in songs on Spotify and using these trends to *predict song popularity*. Factors of acousticness, danceability, energy, duration_ms, instrumentalness, valence, tempo, liveness, loudness, speechiness were considered when looking at popularity. The data is derived from Kaggle's Spotify Multi-Genre Playlists Data, which has data on Spotify songs from seven different genres. Each song in the dataset is ranked on factors such as acousticness, danceability, tempo, liveliness, and more. We trained a Random Forest Classifier model and found the relationship each feature of a song has with popularity.

INTRODUCTION

Define the problem

With the advancement of technology, the world of music consumption has changed dramatically. The internet has made music freely available to anyone in the world, and uploading and sharing music has also become easier than ever. The music streaming market is huge and still continuing to grow. Currently, Spotify is the world's most popular music streaming service, with a market share of 31% and around 165 million subscribers (Mulligan, 2022). As a service dedicated to distributing music, Spotify has algorithms for recommending songs to its users. It recommends popular songs and creates personalized playlists based on what type of music each user listens to. Spotify also has algorithms that score songs based on their 'acousticness', 'danceability', 'energy', 'duration_ms', 'instrumentalness', 'valence', 'tempo', 'liveness', 'loudness', 'speechiness' and 'popularity', which is what we will mainly be observing in this project.

This is important not only to music audiences/enthusiasts, but also to musical artists, such as singers, bands, and songwriters. Looking into this data provides insight on why certain songs are able to generate more interest in communities, which can help artists understand what qualities or modifications to qualities (genre, danceability, energy, loudness, key), they can include or make to their music to attract a wider range of audiences and popularize their efforts. It is also said that "music and identity are inexorably linked, a reflection of our personal history, individuality, and perhaps more importantly, our humanity" (Resler, 2017). Thus, this research can aid understanding in communities and how music preferences can shape people's identities.

Motivation and objectives

For this project, we have one main goal:

- Identify any trends or patterns in features of songs on Spotify that make songs more popular and analyze their relationship with popularity.

As users of Spotify, we were motivated to look into predicting song popularity because music is such a prevalent part of our lives. Most of us listen to music every day during activities like walking, studying, driving, or hanging out with friends. We also spend time curating playlists and exploring new songs that fit our music tastes. With people having unique decisions when choosing music, we wanted to explore the why in what makes certain songs popular. Are there trends in popular songs that make them more well-known? By looking into this data, we can also see how our popularity predictions match our personal taste in music.

Another motivation for looking into song popularity was finding new music to listen to. One of the great things about Spotify is its huge database of music, but this also makes it hard to find new music because there are so many options to look through. We think that looking into song popularity could be a great way for us to find upcoming trends in music and new songs that we typically might not listen to.

METHODOLOGY

Acquisition

From Kaggle, we found 7 datasets, each with information of a different genre. These genres include alternative, rock, metal, pop, indie alternative, blues, and hip hop music. Each of these datasets contain information such as "danceability," "energy," "loudness," and many more. The data was derived from the first 100 songs in each of Spotify's official playlists in 2021 using the Spotify library for Python.

Link: https://www.kaggle.com/datasets/siropo/spotify-multigenre-playlists-data/metadata

Preparation

For the initial stages of EDA, we inspected each of the seven data sets, identified their dimensions for any null or invalid values, and replaced or removed the identified data accordingly. We also checked the types of the columns and made sure all the columns had the proper data type. Lastly, we checked for and removed duplicate rows. All seven datasets had no null-values and the data looked good. There was also no significant number of "0" values for the columns of numerical data (e.g., danceability, energy). There were no duplicate rows nor data type conflict. We used the range and median values from the set of descriptive statistics for each genre to determine whether there were any outliers within the datasets. Overall, we found that this dataset was good.

Since we wanted to analyze features to predict popularity across all songs, we concatenated the seven datasets into one. This was simple because the datasets contained the same headers. The features we chose to extract were the following: acousticness, danceability, energy, duration_ms, instrumentalness, valence, tempo, liveness, loudness, speechiness. We kept most of the data and removed the other columns that could not be quantified into numerical data and used regression on or turned into numbers, such as Track Name or Artist Name.

Model Selection

Our goals for this project are to identify the trend of any feature(s) that affect the popularity score and make predictions based on that trend. The algorithms we considered were logistic regression, decision tree classification, random forest classification, and k-nearest neighbor. Since there is a large range to popularity scores, we chose to split the popularity column into two bins: "popular" and "unpopular". Anything with a popularity score greater than or equal to 50 is "popular" and anything less than 50 is "unpopular". This allowed us to run classification algorithms on our data set, which we believe will be more accurate.

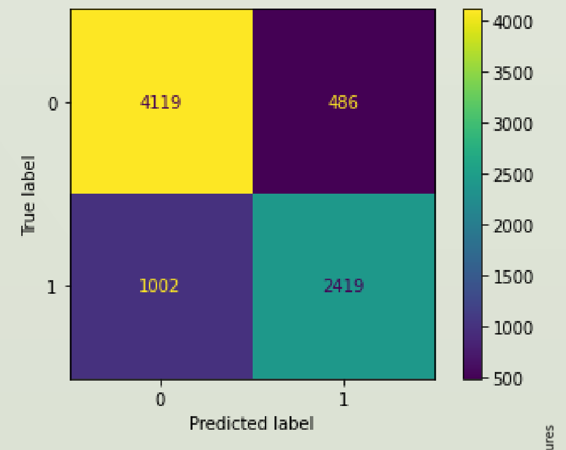
Overall, we think the best model to work with is the random forest classification model. The logistic regression had a low accuracy, so we ruled it out. The decision tree classification had good accuracy, but the random forest was slightly better. Similarly, the K-nearest neighbors model had a lower accuracy compared to random forest. With random forest, we can also easily get the feature importances, which is important for our goal.

For the random forest model, we tested with a number of trees from 600 to 1000. We chose to test large numbers of trees because our dataset has over 26,000 rows. We found that the accuracy was highest with 800 trees. Using 800 trees, we got an accuracy of 81.5%. This was better than decision tree classification and K-nearest neighbors, which were both around 70%.

The features we explored were acousticness, danceability, energy, duration_ms, instrumentalness, valence, tempo, liveness, loudness, and speechiness. The target was popularity (is popular, is not popular). We partitioned the data into a training set (with 70% of data) and test set (with 30% of data) and seeded the data.

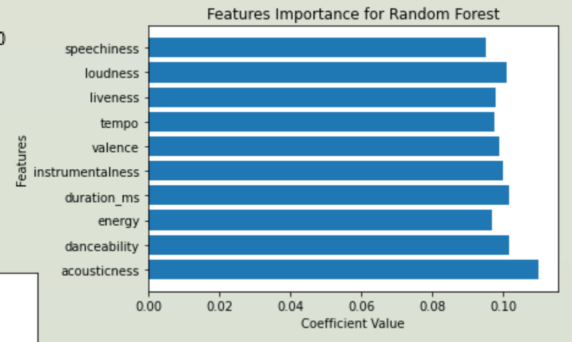
RESULTS & EVALUATION

Using the best model — random forest classification model — for our problem, when we split the popularity into two bins ("unpopular" and "popular") and used 800 trees, the accuracy of the model was around 81.46%. The image below shows the classification report for our model. As it shows, the overall accuracy of about 81% is higher than that of any other models we have experimented with, but the recall is higher for "unpopular" bin, which tells us that it is more likely to accurately identify all positive instances (both true and false positives) for "unpopular". This implies despite the comparatively high overall accuracy of the model, the model is less likely to identify all positive instances for "popular", which somewhat feeds to our initial thought: given how weak correlation between each feature variable and popularity score is, we doubt that the model would not only accurately classify songs as "popular" and "unpopular" but also identify a set of features that greatly influence the popularity of each song.



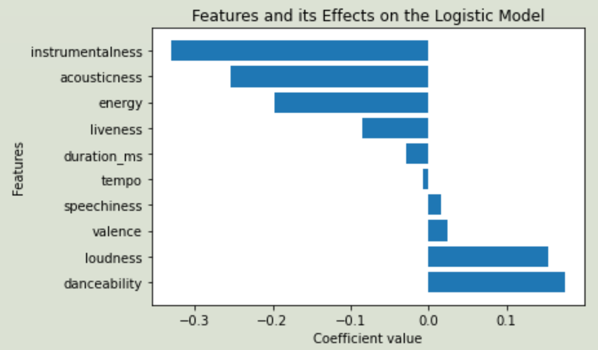
From the confusion matrix on the left, we can see that we have a very high amount of True Positives (4119) and True Negatives (2419) and a very low amount of False Negatives and False Positives. Therefore, our model is doing a good job avoiding Type 1 and Type 2 errors although not entirely. This makes sense since our model has approximately 81% accuracy.

	precision	recall	f1-score	support
unpopular	0.80	0.89	0.85	4605
popular	0.83	0.71	0.76	3421
accuracy			0.81	8026
macro avg	0.82	0.80	0.81	8026
weighted avg	0.82	0.81	0.81	8026

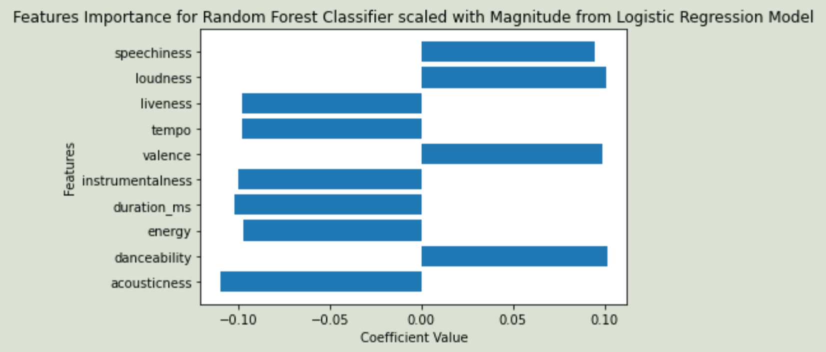


In the feature importances of the random forest model, acoustic-ness has the highest score and speechiness has the lowest score. However, all features have very similar feature importance scores. Also, we also cannot tell if each feature has a positive or negative impact on the popularity score just by looking at the feature importances. In other words, we cannot say for certain that a song with a higher acoustic-ness score will be more popular than a song with a higher speechiness score. For this, we need to look at the feature importances from our logistic regression.

Looking at the importances from the logistic regression, we can see that acoustic-ness also had a big impact on the model, but it actually had a negative impact. This tells us that it likely had a negative impact on popularity in the random forest classification model as well. In the logistic model, danceability had the highest positive impact. It is also high in the random forest model, so danceability may be more important than the other features for popular songs. Ultimately, all features do have very similar feature importance scores, so we believe that popular songs will likely require a combination of all 10 features.



We are using the logistic regression model's coefficient values to determine if the features are going to have a negative impact or a positive impact on the positivity. We are still going to use the coefficient values from our random forest classifier model. However, the signs (+ and -) of the coefficient values will be based on the ones shown in the logistic regression model. We're doing this because the random forest classifier does not have a built-in function to illustrate the impact's magnitude on the popularity. This might affect the accuracy and the results of our investigation.



After looking at the visual on the left, we can conclude that the relationships between the features and the popularity of a song is presented below:

	Features	Impact	Coefficient Value
0	speechiness	positive	0.095091
1	loudness	positive	0.100808
2	valence	positive	0.098899
3	danceability	positive	0.101737
4	liveness	negative	0.097882
5	tempo	negative	0.097604
6	instrumentalness	negative	0.099763
7	duration_ms	negative	0.101650
8	energy	negative	0.096743
9	acousticness	negative	0.109823

Increasing the speechiness, loudness, valence, and danceability is going to impact the popularity of a song positively. We think this is because with these features the song can be more preferable in festivals and concerts as it creates more of an energetic/hype atmosphere and environment for the audience to entertain themselves. Increasing the liveness, tempo, instrumentalness, duration, energy, and acoustic-ness will have a negative impact on popularity. We think that this might be because too strong of instrumentals will not be able to highlight the singer's voice and the meaningful lyrics behind a song, therefore it might not create a considerable impact on the audience.

IMPACTS

With our random forest model (plus a little help from our logistic regression model), we were able to identify that "danceability" and "loudness" had the most positive impact in the popularity of songs on Spotify. However, given how similar the feature importances are, we cannot say the two identified features have an overwhelmingly positive impact on the popularity of songs. Here, we conclude that holistically, there is no apparent trend from a single feature that predicts a song's popularity. However, with the impacts each feature has on a song that we interpreted from the results, this information can still be utilized by artists and producers, especially the one that just recently got into the music industry, to create music with the right level of features in order to raise the odds of the song becoming more popular. We believe our results can aid music companies, indie artists and a lot of other people that are interested in making music since most of them want their music to be heard by the majority of the world.

Due to the fact that we were not able to identify a trend, our identified solution may not have a significant impact. However, there is still one takeaway we can interpret from our failure: popularity of a song is determined by something much more complex than just the influence of a single feature. This information would be more relevant for artists who strive to create popular songs. This means that artists should not focus on single aspects of their work but to create music holistically and stick with their intuition.

CONCLUSIONS

Our objectives of this project were only partially realized. We were able to get a model with a high accuracy and identify the features that had the most impact on popularity. We observed that acoustic-ness may have a negative impact on popularity while danceability and loudness may have a positive impact. But because the features all had very similar importance scores, we cannot assert for certain that this is the trend that cases always follow. In other words, while we were able to see the relationship between each feature to popularity of the songs, we failed to identify any outstanding trend when looking at the features as a whole. This is because a feature might also be influenced by other features as well, and we have not explored dependency of each feature yet. Also, there exists other factors that can influence its popularity as well.

Future Work and Potential Improvements:

Extending from here, our project illustrates that other than the features we have studied in this project, there may be more factors that affect songs' popularity scores that are not listed in our initial dataframe. We would consider looking more into other factors in the near future. Other factors may include but not be limited to:

- Whether or not the artist(s) of the song has/have current name recognition
- Whether or not the artist(s) has/have previous hit records
- The genre(s) the given artist(s) prefer to focus on the majority of the time
- Whether or not the artist(s) collaborated with other well-recognized artist(s)

Given that there are so many different factors that potentially positively or negatively affect the song's popularity score, this serves as an advice to artists to not focus on a single aspect but rather view their work in a more holistic and intuitive manner.

The model also could not explore how different *combinations* of these factors affected popularity. Perhaps a low energy song with high danceability could make a song more popular. If we narrow our search to be genre exclusive, we could potentially get rid of these codependent factors influencing our results.

One thing to consider is that Spotify did not create the categories based on vote, but by their own metrics. It may be entirely possible that the scores for things like "danceability" are inaccurate to begin with. Therefore, in future work, we could consider different features that are based on real people's opinions through surveys about particular songs. Some features that potentially could have been good to know include region, demographics, artist popularity, and on what device it is streamed on.

RELATED WORK

We found two sources that performed similar work. Both of the links below use similar Spotify datasets to predict song popularity based on different song attributes, such as danceability and energy. While they are very similar to the work we want to do, they are not exactly the same because the datasets used in the links below provide more song information compared to the dataset we plan to use. For example, the examples have information on songs' time signatures, tempo, instrumentalness, etc., which is information we do not have.

Links:

- https://towardsdatascience.com/predicting-popularity-on-spotify-when-data-needs-culture-more-than-culture-needs-data-2ed3661f75f1
- https://github.com/MattD82/Predicting-Spotify-Song-Popularity/blob/master/README.md

REFERENCES

Mulligan, Mark. "Music Subscriber Market Shares Q2 2021." *MIDiA Research*, 18 Jan. 2022, https://www.midiaresearch.com/blog/music-subscriber-market-shares-q2-2021.

Resler, Seth. "5 Reasons Why It's Important to Teach Popular Music History." *Jacobs Media Strategies*, 28 July 2017, https://jacobsmedia.com/5-reasons-why-its-important-to-teach-popular-music-history/.