Faculty of Engineering Technology, Campus Group T

Music genre prediction

Jens Coosemans & Michiel Creemers

Introduction

With 23% of all music consumed worldwide coming from streaming services like Spotify and Apple Music, it becomes very important to get to know the listener. One interesting way to improve the listening experience is by providing the best song recommendations, and thus genre classification is necessary.
In the past, some research has been done towards this classification, but this was often done on audio features, like the Mel-Frequency Cepstral coefficients, and these have several issues:

- Require scientific knowledge
- Hard for customer to specify what exactly they are looking for in a song

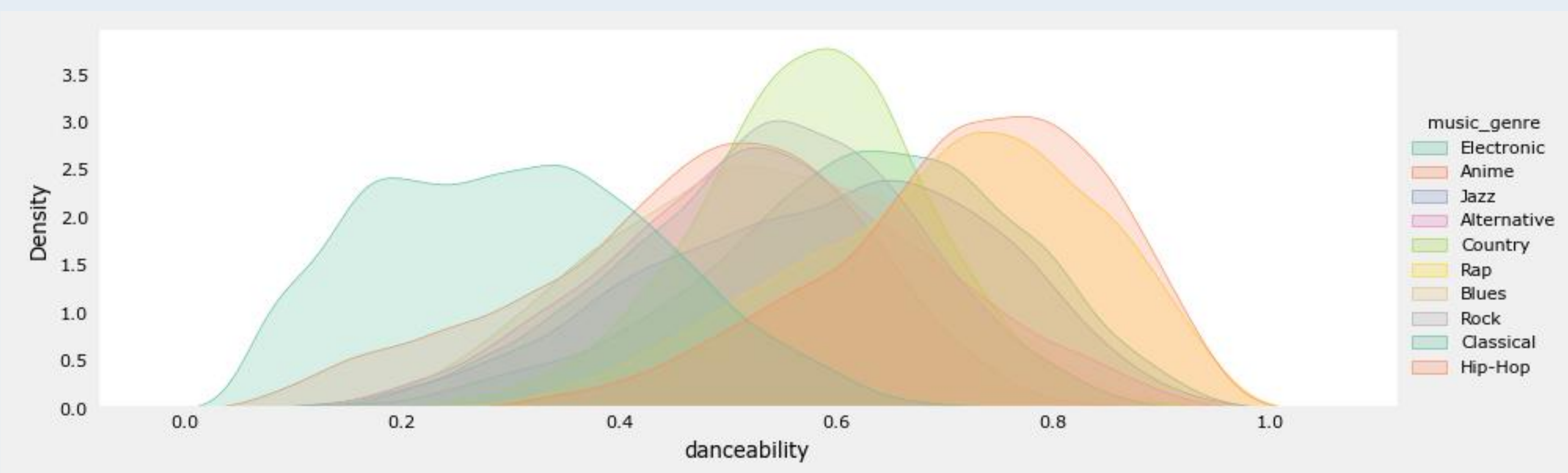
Because of this, we propose a Multi class classification, using the features from the Spotify API

Dataset and Features

Based off the Spotify Audio features

- 50000 entries
- 10 unique genres
- CSV-format
- 17 features

Features: instance id, artist name, track name, popularity, acousticness, danceability, duration ms, energy', instrumentalness, key, liveness, loudness, mode, speechiness, tempo, valence and obtained date
These are easier for most listeners to understand.
The features were also plotted to see which could be dropped, because they weren't relevant.



It turns out that the features instance id, artist name, track name, tempo and obtained date could be dropped.
The features key and mode had to be one hot encoded. The only other required preprocessing was to drop NaN values

Methods

Several classification methods were explored, both from the lecture as more specialized models.
The first approach was by using logistic regression. Since it is a multi class classification, the one vs all approach was used.
The next method from the lab session was the use of a Dense layered Neural Network with the following structure:

Layer(type)	Activation	Output Shape	# Param
Dense	ReLU	512	5632
Dense	ReLU	256	131328
Dense	ReLU	128	32869
Dense	ReLU	64	8256
Dense	Softmax	10	650

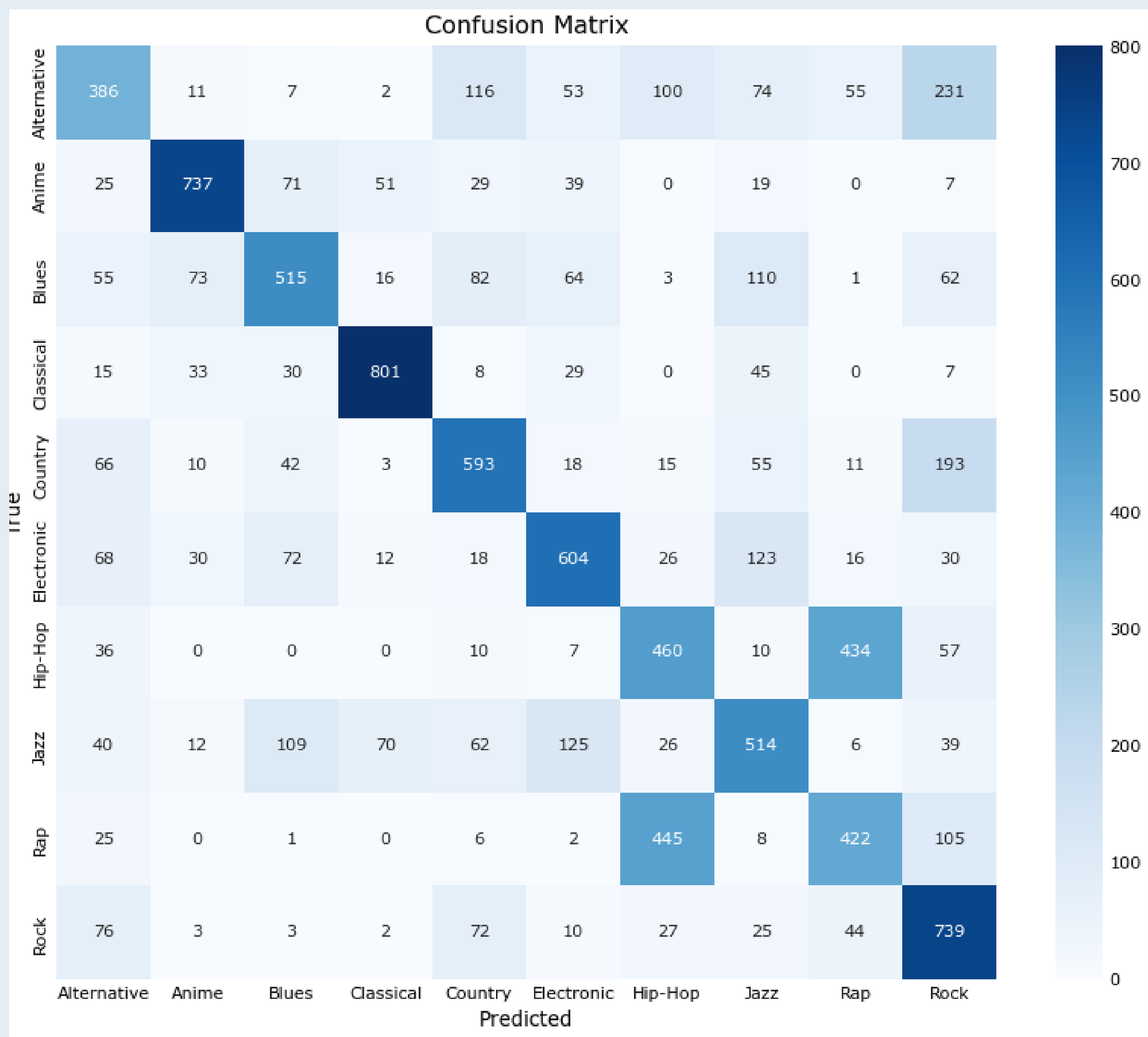
Then finally, specialized methods like Random Forest, Gradient boost and a Grid Search was done to optimize the hyperparameters, but only to a certain extent.

Results/Discussion

The Implementation for the Layered Neural network turned out to be insufficient for the given dataset. Many different combinations were made, but an accuracy of more than 14% was never obtained.
Logistic Regression resulted in an accuracy of around 89% percent with the one vs all approach, but it was not possible to get a prediction through all the classes.
For the more complex models, the results turned out to be way better, they scored the following accuracies

- Random Forest: 55,7%
- Optimized Random Forest: 56.6%
- Gradient Boosting: 58.8%

For the Gradient Boosting the following confusion matrix was made to understand the performance, and the shortcomings of the model.



From this, there are two interesting takeaways. First and foremost, the genres Hip-Hop and rap seem to be very hard to distinguish. This can be explained by the fact that Hip-Hop is more related to the culture around it, while Rap is not per se connected to this broad Hip-Hop culture. In one test, these two genres were put together, which resulted in an accuracy of 67,7%, an increase of nearly 10%.
Another genre that underperformed is Alternative. The main reason for this misclassification can be allocated to the dataset itself. By looking into it, it becomes clear that a lot of subgenres are labeled alternative. For example, songs of the band Linkin Park, which is considered an alternative rock band, were also put in the Alternative genre. This was also the case for metal, and other subgenres.

Conclusion/Future Work

Considering the large number of categories, the result of the classification can be considered ok. In future work, a more specific dataset should be created with more specific genres. Also, a more in-depth Grid Search can be done to optimize the Gradient Boosting since this was to computationally intensive for our hardware.

Another thing that could be done is to combine both the audio features from the Spotify with actual sound snippets from these songs and make a combined model.