

Classifying Music Genres

Tana Liu and Tajhini Brown, CSC 293: Machine Learning, Spring 2025



Introduction

There is over 6,000 genre classifications on Spotify. From psych-rock to minimal techno, and forro to ska, what features define the boundaries of these genres? Our project set out to answer this question and classify songs by these features into genre using random forests and multi-classification.



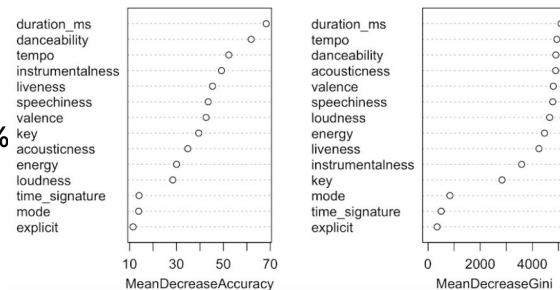
Methods

Using random forests, we analysed the importance of each feature for each genre as well as the overall importance of a feature to the model based on its mean decrease of accuracy and total decrease in node impurity.

We optimised the mtry parameter of the random forest to improve the accuracy, increased the number the number of trees to 100 for stability, and removed the variables that proved to be the least important.

Results

Before our optimisation, our classification accuracy was 23.1% and after it was 24.4%.



J-Rock

- Top 3: danceability (4.53), instrumentalness (4.3), acousticness (3.85)
- Bottom 3: liveness (-1.16), mode (-0.64), tempo (-0.23)

Conclusion

An accuracy rate of 24.4% is dismal but we gather that this may due to the blurring of genre boundaries in recent music history as our data set is from 2022.

Future research would involve using more complex classification techniques like a neural network and a more reliable data set to see if that improves our classification accuracy.