Introduction

We aim to create a model predicting the number of streams a song is likely to receive using a dataset that includes artist information, release details, and audio features. Our goal is to explore how these features influence a song’s popularity and build a regression model for stream prediction.

We want to see if it is possible to see how popular a song will be. If possible, could this be beneficial for the music industry? If you know what a hit will be you could target your marketing strategies, you could optimize automatic playlist creations and thereby enhance user experiences and more.

Materials and methods

We have chosen a dataset of the most streamed songs in 2023. The list doesn’t only contain songs that are released in the year 2023. The list also contains songs that have been released going back to the 1960’s and thereby provides us with a little broader variety of songs.

We utilized a Random Forest Regressor to capture the correlation between a song's characteristics and its streaming figures. We opted for this method due to its resilience to outliers and its capacity to manage non-linear associations.

To optimize the model's performance, we applied Grid Search for hyperparameter tuning.

Literature Cited

Géron, A. (2023). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Datasource: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023/data

**BUILDING A MODEL THAT CAN PREDICT SONG STREAMS**

Results

Feature importance:

A screenshot of a computer

Description automatically generated

Feature tuning and R2 result

A computer screen with white text

Description automatically generated

A graph with a line and a line

Description automatically generated with medium confidence

Conclusion

We found that there is a strong correlation between the number of streams a song gets, and the total amount of playlists its added to.

The model got an R2 score of 0.8 which is okay but further tweaking of the model might led to

For the other features in the dataset not much correlation was found.

Acknowledgements

Using the 'Most Streamed Spotify Songs of 2023' dataset to explore streaming trends and developing a supervised regression model to predict the streams for a song based on selected features.