Project Proposal: Analyzing Spotify Song Attributes for Regression and Classification

Stats 313
Jack Krebsbach and Eli Edwards-Parker

Source: https://www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs

Introduction

This project will use a dataset of nearly 30,000 songs from Spotify, focusing on two main problems: one regression and one classification. By analyzing this dataset, we aim to uncover patterns and relationships within these music samples. The regression problem will delve into how various attributes such as tempo, key, and duration impact a song's popularity. Meanwhile, the classification problem will explore the factors of genre determination, leveraging the audio features provided by Spotify.

Problems to Address

1. Regression Problem: Predicting Song Popularity
2. Classification Problem: Classifying Songs into Genres

Importance and Specific Questions

Regression
Understanding factors influencing song popularity can provide insights into listener preferences and industry trends. What song attributes most significantly predict a song's popularity on Spotify?

Classification
Efficient genre classification aids in better music recommendation and library organization. Can we accurately classify songs into specific genres based on their audio features?

Data Set

Source: Kaggle/Spotify API (Nearly 30,000 Songs)
Observations and Predictors: The dataset includes over 1,000 observations with numerous predictors, such as danceability, energy, key, loudness, and tempo.

Variables

Regression Analysis
- Response Variable: Track Popularity (track_popularity)
- Predictors: Danceability, Energy, Key, Mode, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, Duration

Classification Analysis:
- Response Variable: Playlist Genre (playlist_genre)
- Predictors: Same as above, excluding the response variable used in the regression analysis.

Expected Analyses

Regression Analysis: Linear regression to explore the relationship between song attributes and their popularity.

Classification Analysis: Use of classification algorithms like decision trees, random forests, or support vector machines to classify songs into genres.

References

1. Seon Tae Kim, Joo Hee Oh, Music intelligence: Granular data and prediction of top ten hit songs, Decision Support Systems, Volume 145, 2021, 113535, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2021.113535

2. Ramírez, J., Flores, M.J. Machine learning for music genre: multifaceted review and experimentation with audioset. *J Intell Inf Syst* 55, 469–499 (2020). https://doi.org/10.1007/s10844-019-00582-9

3. Narkhede, Nandkishor & Mathur, Sumit & Bhaskar, Anand. (2022). Machine Learning Techniques for Music Genre Classification..