Nastassia Pukelik

Prof. Jairam

CIS 9660

August 4, 2025

Technical Report: Predicting Valence in Metal Songs Using Regression

Business Problem

This project focused on predicting a song's emotional valence score, which measures how happy or sad it sounds (from 0 to 1). The goal was to build a model that could estimate this score using audio features, helping streaming platforms better understand and recommend emotionally targeted content. I specifically looked at metal songs to explore the emotional diversity within this genre.

Data and Preprocessing

The dataset came from Kaggle: Spotify Tracks Dataset. I filtered it for metal-related genres and used features like danceability, loudness, speechiness, instrumentalness, and tempo. I handled missing data, scaled the numeric features, applied one-hot encoding for key categories, and created a training/testing split (80/20).

Modeling and Results

I trained six models: Linear, Ridge, Lasso, Random Forest, Gradient Boosting, and XGBoost. After comparing metrics like R^2, RMSE, and MAE, XGBoost performed the best:

- R^2 (Train): 0.557
- R^2 (Test): 0.510
- RMSE (Test): 0.137
- MAE (Test): 0.108

XGBoost consistently outperformed the others and had strong cross-validation scores, so I selected it as the final model.

Insights and Takeaways

Danceability was the most important feature positively associated with valence. Instrumentalness and speechiness had negative effects. Heavy metal and metalcore subgenres showed higher valence on average, while black and death metal leaned more negative. Tempo had a slight positive trend, but key and time signature had minimal impact.

Limitations and Next Steps

The R^2 score shows the model explains just over 50 percent of the variance in valence. That's decent, but valence is subjective and affected by lyrics, vocals, and emotional context, which this dataset doesn't capture. Adding sentiment analysis or experimenting with neural networks could improve accuracy in future versions.

Kaggle Dataset link: https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset/data