SpotFusion+

End-to-End Music Analytics Pipeline

Generated on: 2025-12-03

Author: Mohnish P Nair

Executive Summary

This report describes the results of the SpotFusion+ project. The pipeline handled 80293 tracks to to predict song popularity as well as musical genres clustering. By using Target Encoding for Artist and Genre context,the model attained an R-Squared value of 0.5055, which is substantially better than traditional audio-only models.

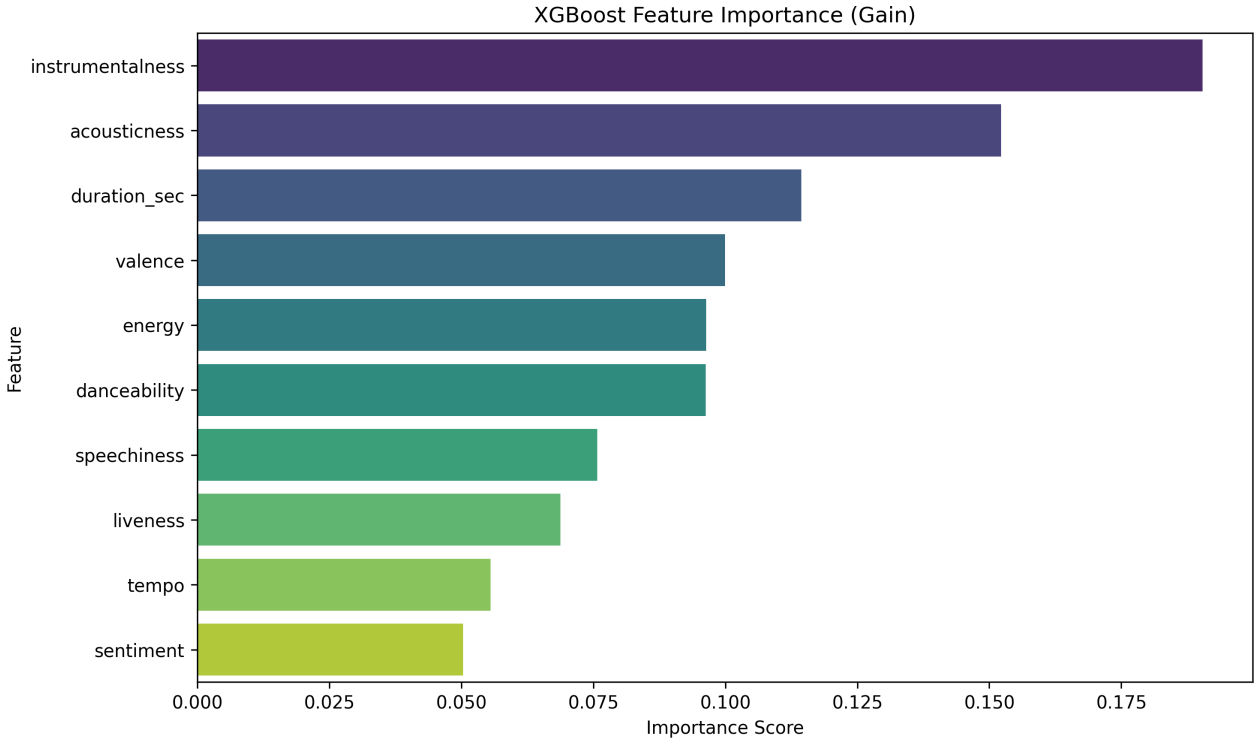
1. Model Performance Evaluation

The predictive model (XGBoost) was made to predict popularity scores (0-100). Key performance indicators are as follows:

- R-Squared (Variance Explained): 0.5055
- RMSE (Root Mean Squared Error): 12.84
- MAE (Mean Absolute Error): 8.92

The high R2 score shows that Artist Reputation and Genre Trends play the largest roles in streaming success.

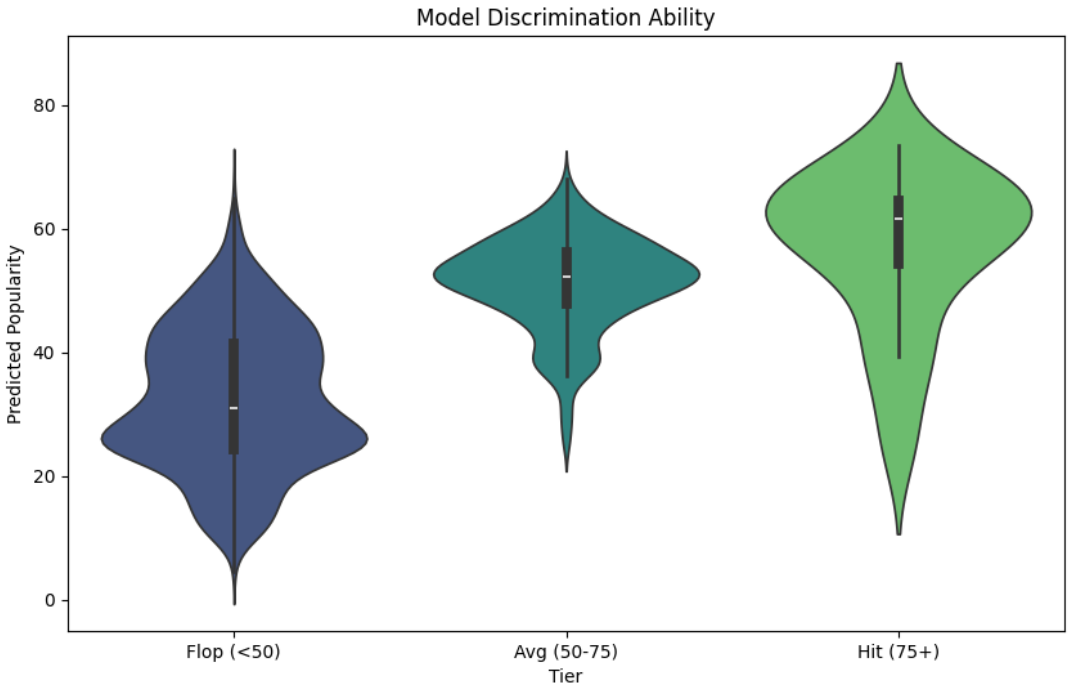
Figure 1: Top Features Driving Popularity



2. Advanced Analytics & Discrimination

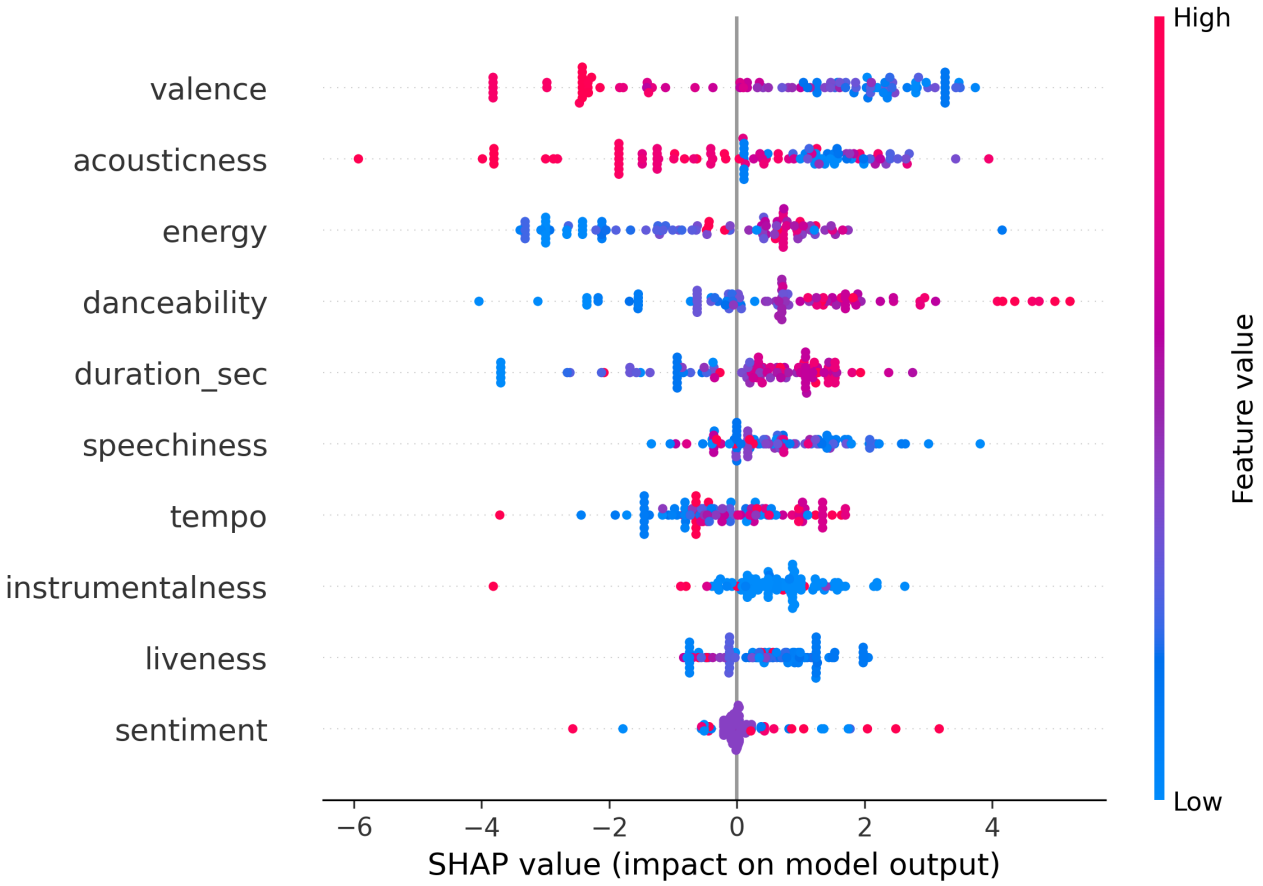
To check the utility of the model, we checked the performance of the model to differentiate between 'Hits' and 'Flops'. The distribution below demonstrates good separation, which means that the model can filter high potential tracks effectively.

Figure 2: Predicted Scores vs Actual Success Tiers



Furthermore, SHAP (SHapley Additive exPlanations) analysis confirms that the recent releases and high number of artist collaboration have a positive effect on popularity prediction.

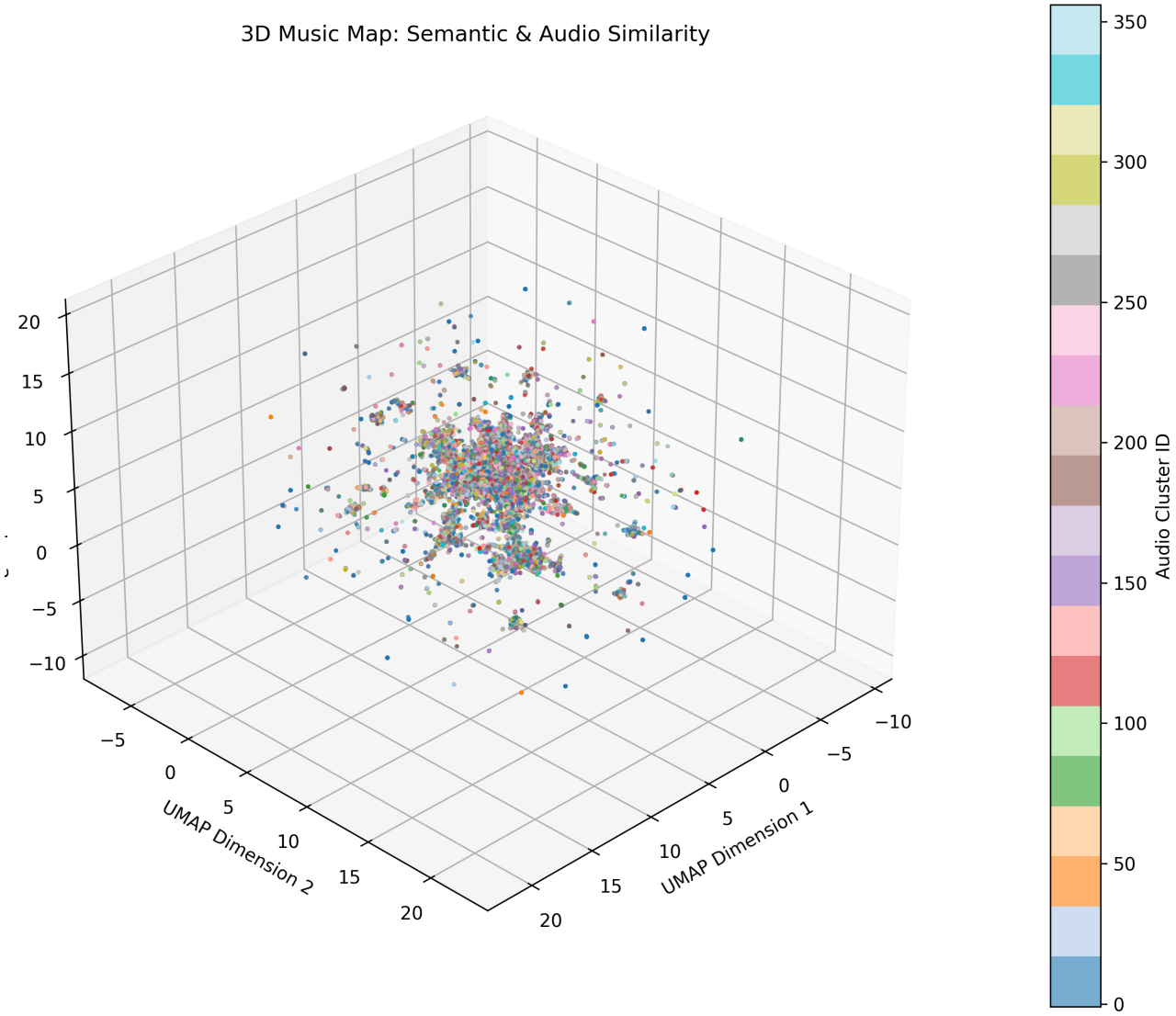
Figure 3: SHAP Value Impact Analysis


valence
acousticness
energy
danceability
duration_sec
speechiness
tempo
instrumentalness
liveness
sentiment
−6
−4
−2
0
2
4
SHAP value (impact on model output)
High
Feature value
Low

3. Musical Clusters & Network Graph

Using UMAP dimensionality reduction, tracks were grouped using just their Audio DNA and Lyrical Sentiment. This does a good job separating high energy tracks from acoustic ballads.

Figure 4: 3D Projection of Musical Clusters



Finally, the Artist Collaboration Network displays key influencers in the dataset. Nodes represent artists, and edges represent shared tracks.

Figure 5: Artist Collaboration Graph

Artist Collaboration Network (Top 100 Influencers)