Sonic Signature: Computational Music Analysis

From Machine Learning Classification to Audio Fingerprinting

AARES Project Report

February 11, 2026

Contents

1 Introduction

This report details the implementation and results of the "Sonic Signature" project, a comprehensive study applying data science and signal processing to music. The work is divided into:

1. **Part 1**: Metadata analysis using Spotify audio features for genre classification, popularity regression, and recommendation.
2. **Part 2**: Signal processing to build a "Shazam-like" audio fingerprinting system and uncertainty quantification via Conformal Prediction.

2 Part 1: Spotify Machine Learning Analysis

2.1 1.1 Dataset and Features

The analysis leverages the `spotify_dataset_train.csv` dataset. Key audio features include:

Feature	**Description**
`acousticness`	Confidence measure (0.0-1.0) of whether the track is acoustic.
`danceability`	Suitability for dancing based on tempo, rhythm, and beat strength.
`energy`	Perceptual measure of intensity and activity.
`instrumentalness`	Likelihood the track contains no vocals.
`valence`	Musical positiveness (happy/cheerful vs. sad/depressed).
`tempo`	Estimated tempo in BPM.

Table 1: Feature Dictionary

Preprocessing Steps:

- Extracted `year` from `release_date` to capture temporal trends.
- Imputed missing numerical values with the **median** and categorical values with the **mode**.
- Standardized numerical features ($\mu = 0, \sigma = 1$) using `StandardScaler` and one-hot encoded categorical features like `key` and `mode`.

2.2 1.2 Genre Classification

We trained a **Random Forest Classifier** (100 estimators, balanced weights) on the processed data.

2.2.1 Results

- **Cross-Validation F1 Micro Score**: **0.4396** ($\pm$ 0.0132)

While the model performs better than random guessing, the F1 score of ≈ 0.44 highlights the inherent ambiguity in genre boundaries based solely on high-level audio descriptors.

Submission: Predictions for the test set were generated and saved to `outputs/submission.csv`.

2.3 1.3 Popularity Prediction

A **Random Forest Regressor** was used to analyze the `spotify_dataset_subset.csv`.

- **Mean Squared Error (MSE)**: **517.99**
- **R-squared (R^2)**: **0.2126**

The relatively low R^2 (21%) suggests that audio features explain only a small fraction of a song's popularity, confirming that external marketing and artist fame play a more dominant role.

2.4 1.4 Recommendation System

A content-based recommender using **Cosine Similarity** was built. It effectively clusters songs with similar vectors (e.g., retrieving high-energy EDM tracks when queried with an electronic song).

3 Part 2: Audio Fingerprinting

3.1 2.1 Methodology

We implemented a robust identification system based on spectral peak hashing.

- **Spectrogram**: Computed via STFT ($n\_fft = 2048, hop = 512$) at a sampling rate of **3000 Hz**.
- **Constellation Map**: Local maxima were extracted from the spectrogram.
- **Hashing**: Pairs of peaks (Anchor, Target) within a target zone were hashed using:

$$\text{Hash} = f_{anchor} \times 10^6 + f_{target} \times 10^3 + (t_{target} - t_{anchor})$$

3.2 2.2 Search Results

A database was built from the `songs/` directory containing 13 tracks. When querying with an excerpt from *"Carmen Prelude"*, the system successfully identified it.

Top Matches:

1. **Bizet, Georges - Carmen Prelude** (Correct Match)
2. Debussy - Suite Bergamasque - 2. Menuet
3. Debussy - Suite Bergamasque - 4. Passepied

The correct song appeared at the top, validating the temporal alignment algorithm.

4 Bonus: Conformal Prediction

To address the uncertainty in our genre classifier (Part 1), we applied Inductive Conformal Prediction using a KNN classifier ($k = 5$) on a calibration set of 1500 samples.

4.1 Empirical Results

For a randomly selected test point, we computed prediction sets at different significance levels (ϵ).

Error Rate (ϵ)	**Confidence**	**Prediction Set**
0.05	95%	`['blues', 'classical', 'country', 'dance', 'disco', 'edm', 'electro', 'folk', 'hip hop', 'jazz', 'latin', 'pop', 'r&b', 'rap', 'rock', 'soul']` (Size: 16)
0.10	90%	`['classical', 'country', 'dance', 'edm', 'electro', 'folk', 'hip hop', 'jazz', 'pop', 'r&b', 'rap', 'soul']` (Size: 12)
0.20	80%	`['country', 'electro', 'folk', 'hip hop', 'pop', 'r&b', 'rap', 'soul']` (Size: 8)

Table 2: Conformal Prediction Sets

Interpretation: To achieve 95% confidence that the true genre is included, the model must return a large set (16 genres), reflecting the high uncertainty and class overlap. Relaxing the requirement to 80% allows for a much tighter set (8 genres).

5 Conclusion

The project demonstrated the strengths and limitations of computational music analysis:

- **Metadata**: While useful for recommendation, high-level features alone are insufficient for precise popularity prediction ($R^2 \approx 0.21$) or fine-grained classification.
- **Signal Processing**: The fingerprinting algorithm demonstrated high robustness, correctly identifying songs despite the simplified sampling parameters.
- **Uncertainty**: Conformal prediction provided a rigorous framework to quantify the classifier's ambiguity, essential for building trust in AI music systems.