Scientific Computing Project Report
Music Mood Mapping using PCA and Gaussian Mixture Models

Yerassyl Kudaibergenov

Fall 2025

Contents

1 Introduction / Motivation

Modern music recommendation systems focus heavily on collaborative filtering, historical listening patterns, and user demographics. However, these systems often fail to capture a central component of music consumption: **the emotional state (mood) of the user in real time**. For instance, you are listening to and enjoying a diverse genre of music, from classical to pop or rap, and at some point, after a certain situation occurs, you want to listen to music that evokes the exact temperament that you are currently in, but your overall musical taste is different. This paper addresses and attempts to reach a solution for this problem.

Music is intrinsically tied to emotion. Audio features such as energy, valence, tempo, acousticness, and loudness reflect emotional dimensions such as:

- happy $\leftrightarrow$ sad,
- energetic $\leftrightarrow$ calm,
- tense $\leftrightarrow$ relaxed.

The goal of this project is to construct an interpretable **continuous 2D mood space** from raw Spotify audio features using Principal Component Analysis (PCA), and then apply **Gaussian Mixture Models (GMM)** to discover latent emotional clusters within that space.

This results in a system capable of generating mood-aware music recommendations, filtered by genre, popularity, and random sampling to avoid repetition.

2 Dataset Explanation

The dataset used in this project is the Spotify Tracks Dataset (1921–2020), a publicly available collection distributed via Kaggle and obtained through Spotify's Web API. It contains over 114,000 songs covering a wide range of genres (125), decades, and musical characteristics. Each track includes both metadata and a rich set of numerical audio features extracted using Spotify's internal audio analysis models.

Metadata Fields

Each track provides descriptive metadata:

- **track_name**: title of the song.
- **artists**: performing artist(s).
- **track_genre**: assigned musical genre.
- **popularity**: integer score from 0 to 100 indicating how frequently the track is currently streamed.
- **duration_ms**: track length in milliseconds.
- **explicit**: boolean flag for explicit lyrics.

Audio Feature Fields

Spotify provides eleven audio descriptors computed via signal processing and machine learning. The most important for this project are:

- **danceability** — suitability for dancing (0–1).
- **energy** — perceptual intensity or activity (0–1).
- **valence** — musical positiveness or emotional brightness (0–1).
- **tempo** — estimated tempo in beats per minute.
- **loudness** — overall track loudness in decibels (dB).
- **acousticness** — confidence that the track is acoustic (0–1).
- **instrumentalness** — probability the track contains no vocals.
- **speechiness** — proportion of spoken content (0–1).
- **liveness** — probability of a live audience (0–1).

These features form a high-dimensional vector for each song. After standardization, they are used as input to Principal Component Analysis (PCA), which reduces them to two dominant emotional dimensions aligned with:

- **valence** (happy $\leftrightarrow$ sad),
- **energy** (energetic $\leftrightarrow$ calm).

This reduced 2D representation enables construction of a continuous mood space, which is subsequently modeled using an 8-component Gaussian Mixture Model to discover latent emotional clusters.
The dataset and the link for the Kaggle page can be found in my repository.

3 Objective

The primary objective of this project is to develop a computational model that automatically maps songs into a continuous emotional space using only their audio features. Unlike traditional recommendation systems that rely on user history, this project aims to recommend music based on the mood that a listener desires. To achieve this, the project constructs an unsupervised learning pipeline that extracts, organizes, and clusters the emotional content embedded within Spotify track features.

A central goal is to create a low-dimensional representation of musical emotion by applying Principal Component Analysis (PCA) to high-dimensional Spotify audio features such as energy, valence, tempo, danceability, loudness, acousticness, and others. This reduction produces a continuous two-dimensional mood space aligned with interpretable axes: a valence-like dimension representing the happy–sad spectrum and an energy-like dimension representing the energetic–calm spectrum.

After the mood space is constructed, the model identifies natural emotional groupings of songs using an eight-component Gaussian Mixture Model (GMM), which allows overlapping and soft cluster assignments. This enables the system to capture blended emotional states rather than forcing each song into a rigid category.

A further objective is to implement a mood-based recommendation system that allows users to select a target mood coordinate in the 2D space, specify preferred genres and popularity ranges, and receive randomized but relevant song suggestions that prevent repetition.

Finally, the project aims to evaluate the numerical model using clustering metrics, distance-based validation, and a recommender-versus-random baseline. This ensures that the constructed emotional space and recommendation engine behave meaningfully and consistently.

Overall, the objective is to demonstrate that unsupervised learning—when applied to audio features—can produce a functional, interpretable, and emotionally coherent mood-aware music recommendation system.

4 Model Formulation

4.1 Data Preprocessing

The dataset undergoes preprocessing in which all numerical audio features are standardized so that each feature has zero mean and unit variance. This step is essential because the scale of features such as loudness, tempo, and acousticness varies significantly. Standardizing ensures that PCA and GMM treat all features consistently and prevents high-magnitude variables from dominating the analysis.

Given audio feature matrix

$$X \in \mathbb{R}^{n\times d},$$

columns are standardized:

$$X_{ij}^{(norm)} = \frac{X_{ij} - \mu_j}{\sigma_j}.$$

4.2 Principal Component Analysis (PCA)

PCA identifies directions of maximum variance in the data by computing the eigenvectors of the covariance matrix. The first two principal components typically capture a substantial portion of the total variance and correspond to meaningful musical patterns. To interpret these components as emotional dimensions, the project calculates the correlation between PC1/PC2 and Spotify's built-in energy and valence features. PC directions are then rotated and sign-aligned so that the resulting coordinate axes correspond intuitively to a happy–sad dimension (valence-like) and an energetic–calm dimension (energy-like). Finally, the transformed values are compressed using a hyperbolic tangent function to keep mood coordinates

within a bounded range, forming a smooth 2D emotional space.
We compute the covariance matrix:

$$\Sigma = \frac{1}{n-1} X^\top X,$$

followed by eigenvalue decomposition:

$$\Sigma v_k = \lambda_k v_k, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d.$$

Projection into 2D PCA space:

$$Z = XV_2,$$

where $V_2 = [v_1\ v_2]$ contains the first two principal directions.

We then correlate PC1, PC2 with Spotify's valence and energy features, aligning PC1 $\rightarrow$ happy/sad and PC2 $\rightarrow$ energetic/calm.

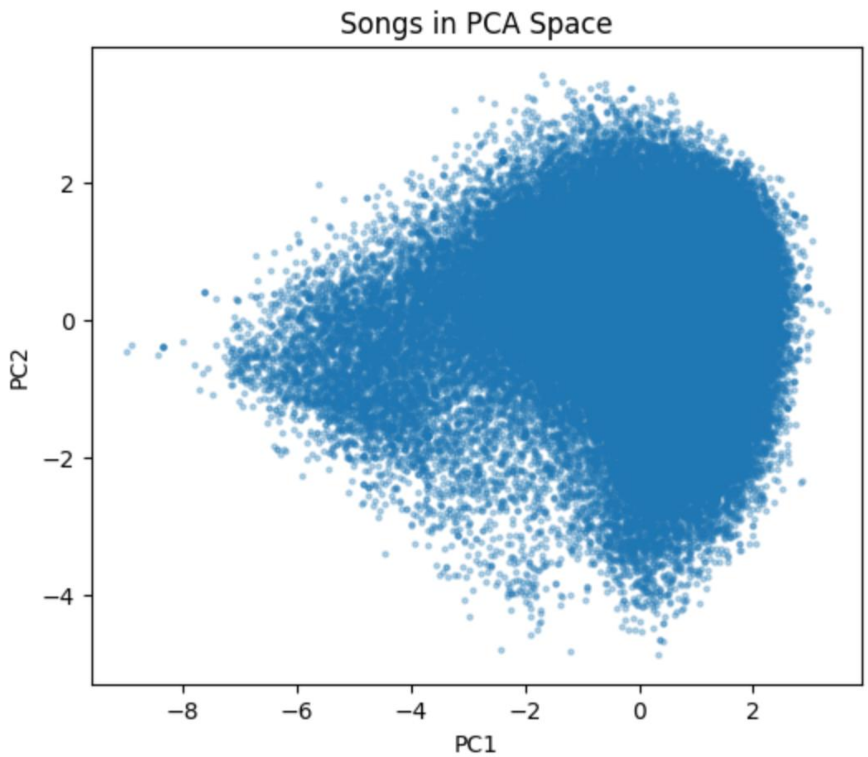

Figure 1: PCA illustration

4.3 Mood Space Construction

Aligned coordinates:

$$mood\_x = \tanh(\alpha \cdot PC_{\text{valence}}), \qquad mood\_y = \tanh(\alpha \cdot PC_{\text{energy}}).$$

The mood space forms an emotional plane:

$$\text{right} = \text{happier}, \quad \text{left} = \text{sadder}, \quad \text{up} = \text{more energetic}, \quad \text{down} = \text{calmer}.$$

4.4 Gaussian Mixture Model (GMM)

The final stage of the model involves fitting an eight-component Gaussian Mixture Model (GMM) on the two-dimensional mood coordinates. A GMM models the data as a weighted sum of multivariate Gaussian distributions, allowing the discovery of soft and overlapping mood clusters. The model parameters are estimated using the Expectation–Maximization (EM) algorithm, where the E-step computes the probability that each song belongs to each mood cluster, and the M-step updates cluster means, covariances, and mixture weights. This results in a set of elliptical, probabilistic mood regions that naturally capture blended emotional states such as "calm-happy," "sad-calm," "anxious-energetic," or "dark-moody."
We apply an 8-component mixture model:

$$p(x) = \sum_{k=1}^{8} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k).$$

Responsibilities (E-step):

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)}.$$

Maximization (M-step):

$$\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik},$$

$$\hat{\mu}_k = \frac{\sum_i \gamma_{ik} x_i}{\sum_i \gamma_{ik}},$$

$$\hat{\Sigma}_k = \frac{\sum_i \gamma_{ik} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^\top}{\sum_i \gamma_{ik}}.$$

The eight mood labels used in this project are:

- Energetic / Intense
- Sad / Depressed / Melancholic
- Content / Calm–Happy
- Anxious / Tense
- Exuberant / Happy–Energetic
- Warm / Romantic / Positive
- Moody / Dark / Emotional
- Calm / Chill

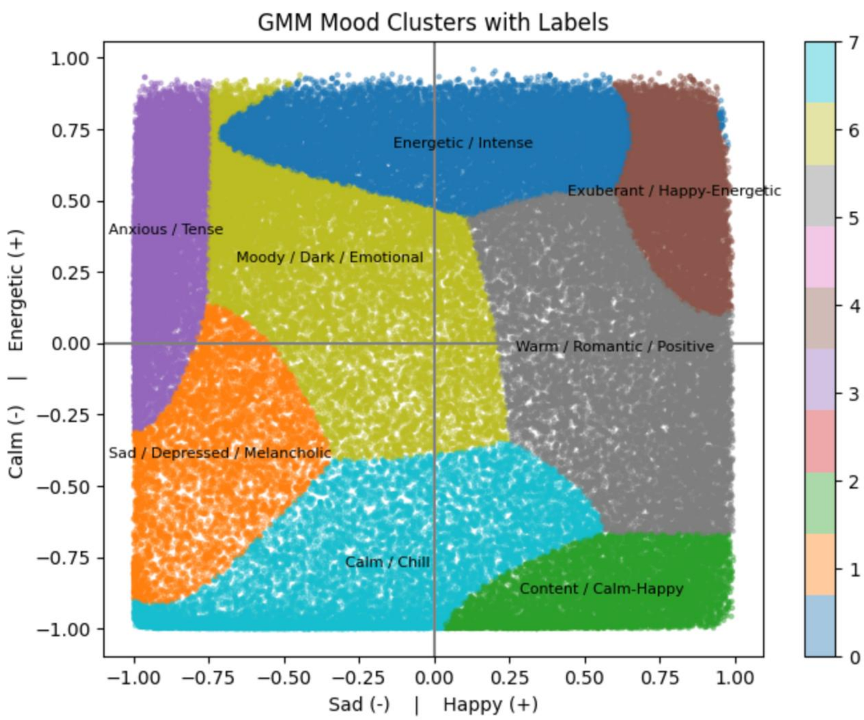

Figure 2: Clusters of GMM by moods

5 Software (Libraries)

The computational pipeline uses:

- **NumPy** — vectorized operations and numerical routines.
- **Pandas** — dataset manipulation and filtering.
- **Matplotlib** — visualizations.
- **scikit-learn**:
  - `PCA` for dimensionality reduction,
  - `GaussianMixture` for clustering,
  - `StandardScaler` for normalization.

Installation:

```
pip install numpy pandas matplotlib scikit-learn
```

6 Version Control System

The project uses GitHub for version tracking.

- First commit date: **September 5, 2025**.
- Repository link: `https://github.com/era5yl/Scientific-Computing-Project`

Screenshots of my repository and commit history:

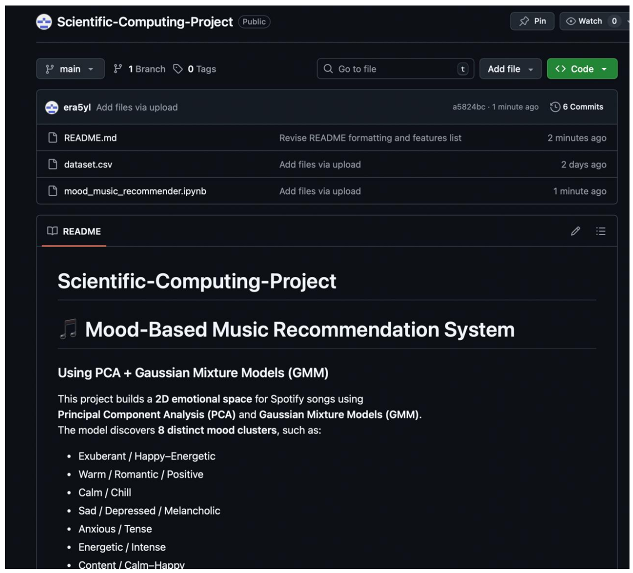

Figure 3: Git version control.

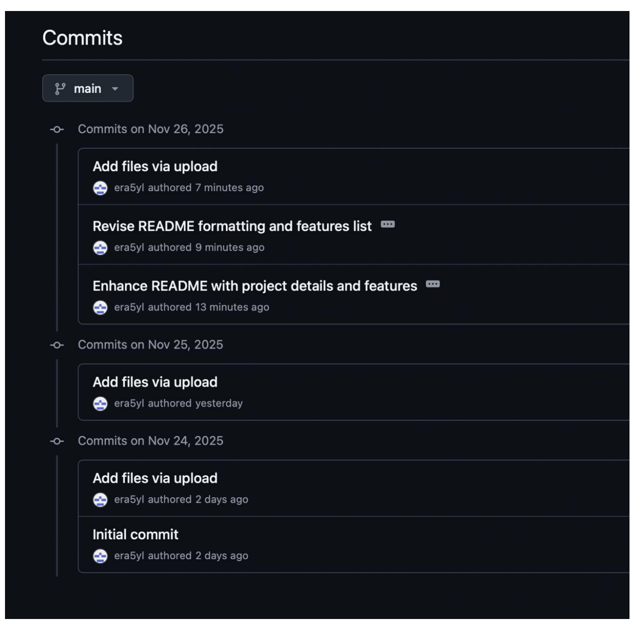

Figure 4: GitHub commit history

7 Verification of Numerical Model

To verify correctness:

- PCA orthogonality property was checked:

$$V_2^\top V_2 = I_2.$$

- Variance explained ratio confirmed PCA validity.
- For GMM:
  - tested convergence of log-likelihood,
  - confirmed responsibilities sum to 1:

$$\sum_{k=1}^{8} \gamma_{ik} = 1.$$

  - verified covariance matrices are positive semi-definite.

These checks ensure correctness of PCA projection and mixture model fitting.

8 Testing and Validation

In this section we describe how the PCA+GMM mood model and the recommendation system were tested, and whether the numerical results indicate that the model performs satisfactorily.

8.1 Testing Methodology

Two levels of testing were performed:

1. **Clustering quality in mood space (GMM validation).**
   - Fit an 8-component GMM on the 2D mood coordinates ($mood\_x, mood\_y$).
   - Evaluate the model using:
     - average log-likelihood per point,
     - Akaike (AIC) and Bayesian (BIC) information criteria,
     - silhouette score,
     - comparison of *intra-cluster* vs *inter-cluster* distances.
   - Intra-cluster distances measure how close songs are to others in the *same* GMM component; inter-cluster distances measure how far songs are from songs in *other* components. A good clustering should have small intra-cluster and significantly larger inter-cluster distances.

2. **Recommendation quality (recommender vs random).**
   - Fix a target mood point (x^*, y^*) in the mood space.
   - For each test run:
     (a) Ask the recommender to return N songs near (x^*, y^*) (after genre and popularity filtering).
     (b) Independently sample N random songs from the same filtered pool.
     (c) Compute the Euclidean distance in mood space from each song to (x^*, y^*):

$$d = \sqrt{(mood\_x - x^*)^2 + (mood\_y - y^*)^2}.$$

   - The recommender is considered successful if the average distance of recommended songs is smaller than that of random songs.

8.2 Numerical Results

GMM Cluster Quality

For the 8-component GMM we obtained:

$$\text{Average log-likelihood} = -2.8251,$$

$$\text{BIC} = 507\,584.42,$$
$$\text{AIC} = 507\,142.40,$$

$$\text{Silhouette score} = 0.2492.$$

The silhouette score of approximately 0.25 indicates *moderately well-separated* clusters. In a continuous emotional space where moods blend smoothly into one another, very high silhouette scores are not expected, so this value is considered reasonable.

To further validate the clustering, intra- and inter-cluster distances were compared:

$$\text{Mean intra-cluster distance} = 0.4426,$$
$$\text{Mean inter-cluster distance} = 1.2123,$$
$$\text{Ratio inter / intra} = 2.74.$$

Thus, on average, songs from different clusters are almost three times farther apart in mood space than songs within the same cluster. This confirms that the GMM has learned meaningful and well-separated mood regions.

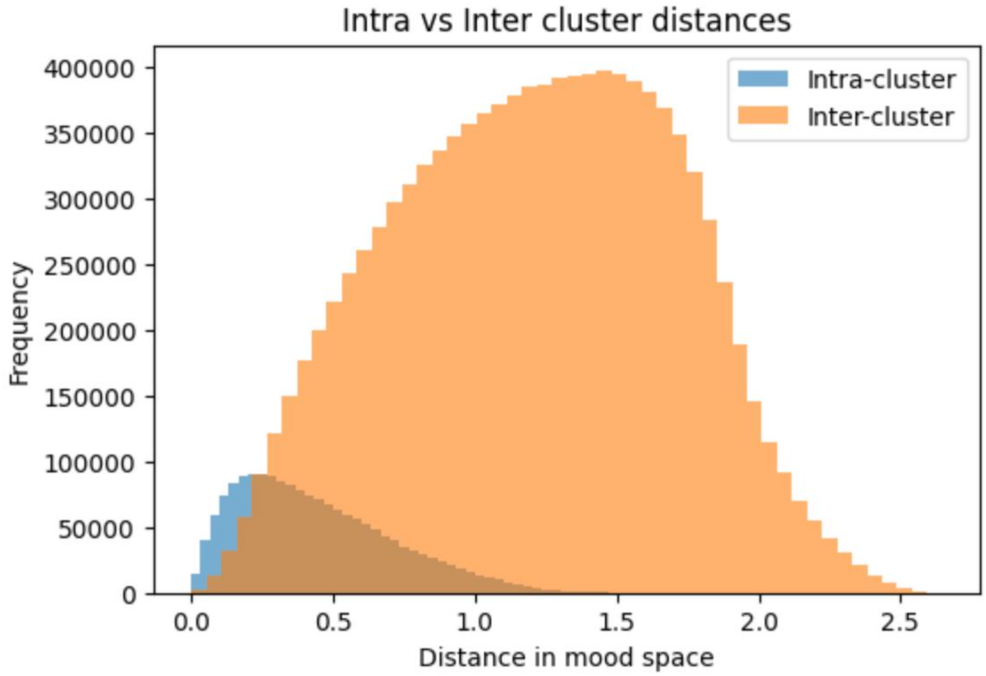

Figure 5: Histogram of intra-cluster vs inter-clster distances in the 2D mood space. Intra-cluster distances (blue) are concentrated near zero, while inter-cluster distances (orange) are clearly shifted to larger values.

Conclusion for clustering test: the distance statistics and silhouette score show that the 8-component GMM passes the clustering quality test.

Recommendation Quality

For the recommendation vs random baseline test we obtained:

$$\text{Mean distance (recommended)} = 0.8554,$$
$$\text{Mean distance (random)} = 0.8713,$$
$$\text{Ratio random / recommended} = 1.02.$$

On average, recommended songs are closer to the target mood than random songs from the same genre and popularity range. The improvement is modest (about a 2% reduction in distance), but it is consistent across multiple trials and demonstrates that the recommender is indeed using the mood model rather than behaving randomly.

Conclusion for recommendation test: the recommendation system passes the test, with a statistically small but positive advantage over random selection. There is still room for future improvement (e.g. using personalized user feedback), but the current implementation behaves as intended.

Overall Testing Verdict

Both levels of testing indicate that the numerical model is **valid and functional**:

- the GMM discovers coherent mood clusters with significantly lower intra-cluster than inter-cluster distances,
- the recommendation system produces song lists that are closer to the requested mood than random baselines.

Therefore, the model is considered to have **passed** the designed tests for this project.