HYBRID- (MOOD BASED)

"Music Recommendation System"

Team Details

Team Name: MUSIFY
Section: B + A

Team Members:

1. Kabeer (Team Leader)
2. Sanskriti
3. Saptak Banerjee
4. Vishudh Goel
5. Naman Kajla
6. Advith Sharma

Team Leader Email:
Kabeer.24bcs10119@sst.scaler.com

PROJECT OVERVIEW

1. Problem Background & Motivation

Music today is no longer consumed randomly; it has become deeply intertwined with human emotions, routines, and mental states. People use music to regulate mood, boost focus, relieve stress, celebrate moments, or find emotional comfort. Despite massive song libraries and sophisticated algorithms, most music streaming platforms still struggle to understand *why* a user wants a particular song at a specific moment.

Current recommendation systems rely heavily on **historical listening behavior**, genre preferences, or popularity signals. While effective to some extent, these systems often fall into repetitive loops, recommending similar genres again and again. This becomes frustrating when a user's **current emotional need** differs from their past behavior. For example, a student who usually listens to energetic pop may want calm instrumental music while studying late at night, or nostalgic melodies during moments of emotional vulnerability.

This gap between **momentary mood** and **historical preference** leads to frequent skipping, reduced engagement, and eventual dissatisfaction. From a business perspective, irrelevant recommendations reduce listening time, playlist creation, and long-term retention, directly impacting revenue.

2. Core Problem Definition

The core problem addressed in this project is:

How can we recommend emotionally relevant music when mood is subjective, dynamic, and unlabeled, and when historical user data may be unavailable or insufficient?

What we aim to discover and recommend:

- **Discover latent emotional structures in music using audio features**
- **Segment songs into mood-based clusters**
- **Recommend songs that match a user's current mood, while still respecting general popularity and preference signals**

Why supervised learning is not suitable?

- **The dataset does not contain explicit mood labels**
- **Emotional perception of music is subjective and varies across users**
- **There is no single "correct" label for a song's mood**
- **Recommendation is about** *ranking and discovery*, **not classification**

Hence, unsupervised learning + recommendation systems is the most appropriate approach.

3. Dataset Description

3.1 Dataset Source

- **Dataset Name:** Spotify Tracks Dataset
- **Source:** Kaggle
- **Link:** https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset
- **Public Dataset:** Yes

3.2 Dataset Overview

- **Initial size: 114,000 tracks**
- **Final size (after cleaning): 89,741 tracks**
- **Number of features: 21**
- **Data types:**
    - **Numerical audio features (float & int)**
    - **Categorical metadata (artists, genre)**
    - **Boolean explicit flag**

3.3 Key Audio Features Used

- **Valence** – positivity of emotion
- **Energy** – intensity and activity
- **Danceability** – rhythm and beat suitability
- **Acousticness** – acoustic vs electronic
- **Tempo** – speed of music
- **Loudness** – perceived intensity
- **Instrumentalness** – absence of vocals
- **Mode** – major/minor tonality

These features are strongly linked to **human emotional perception of music**, making them ideal for mood-based clustering.

4. Data Preprocessing & Exploration

Before applying any machine learning techniques, **extensive preprocessing and exploratory analysis were carried out to ensure data quality, consistency, and suitability for unsupervised learning**. (*Since clustering is highly sensitive to noise and scale, this step was crucial for obtaining meaningful emotional groupings)*.

4.1 Duplicate Handling

During initial exploration, the dataset was found to contain multiple entries for the same song, identified by duplicate track_id values.

- **A total of 24,259 duplicate track_ids were identified.**
- **These duplicates did not represent identical rows but rather repeated tracks with varying popularity values.**
- **To retain the most relevant version of each song, the instance with the highest popularity score was preserved, as it best represents user preference and real-world relevance.**
- **After this step:**
  - **No duplicate rows remained in the dataset**
  - **Each track_id uniquely identified a song**

This ensured data integrity while preserving the most impactful version of each track.

4.2 Feature Engineering

Several transformations were applied to improve interpretability and modeling effectiveness:

- **Duration Conversion:**
  Track duration was converted from milliseconds to seconds for better human interpretability and analytical convenience.
- **Removal of Redundant Columns:**
  Columns not directly contributing to emotional understanding (e.g., identifiers and raw duration after conversion) were removed to reduce noise.
- **Selection of Emotionally Relevant Features:**

  Only intrinsic audio features strongly linked to human emotion were retained,

including features like Valence , Energy , Danceability, Acousticness, Tempo , Loudness, Instrumentalness , Mode.

This ensured that clustering was driven purely by musical emotion, not metadata bias.

4.3 Feature Scaling

The selected features existed on very different numerical scales:

- Tempo and loudness have large numeric ranges
- Valence and danceability lie between 0 and 1

To prevent features with larger magnitudes from dominating the clustering process, **StandardScaler** was applied to normalize all features to zero mean and unit variance.

5. Methodology & Model Choice

5.1 Why Gaussian Mixture Models (GMM)?

Gaussian Mixture Models were chosen over traditional clustering algorithms such as K-Means due to the **nature of emotional data**.

The key reasons include:

- **Overlapping Emotions:**
  Music emotions are not mutually exclusive. A song can feel both calm and sad, or energetic and happy. GMM supports this through probabilistic assignments.
- **Soft Clustering:**
  Unlike K-Means, which assigns each song strictly to one cluster, GMM models the probability of belonging to multiple clusters, making it more realistic.
- **Flexible Cluster Shapes:**
  GMM with full covariance matrices can model clusters of varying size, orientation, and density, which is essential for complex emotional spaces.
- **Continuous Emotional Modeling:**
  Emotional perception exists on a continuum rather than fixed boundaries. GMM aligns naturally with this assumption.

Thus, GMM was well-suited for discovering latent emotional structures in music.

5.2 Model Selection Using BIC & AIC

Selecting the optimal number of clusters is critical in unsupervised learning. To achieve this, we used two standard statistical criteria:

- **BIC (Bayesian Information Criterion)**
- **AIC (Akaike Information Criterion)**

- **Procedure:**

- GMMs were trained for **k = 2 to 10 clusters**
- BIC and AIC scores were computed for each model
- Lower scores indicate a better balance between model fit and complexity

- **Result:**

Both BIC and AIC curves showed a clear minimum around **k = 6**, indicating that six clusters best represent the underlying emotional structure of the dataset without overfitting.

Hence, **six mood clusters** were selected for further analysis.

6. Experiments & Results

6.1 Mood Clusters Discovered

Using GMM with six components, the model discovered the following latent emotional clusters:

Cluster	Interpreted Mood
0	Sleep / Chill
1	Energetic / Party
2	Sad
3	Happy / Upbeat/ Dance
4	Calm / Focus
5	Emotional/Heartbreak

Each cluster exhibited distinct audio characteristics:

- **High energy and tempo** → Energetic / Party cluster
- **High acousticness and low loudness** → Calm and Focus clusters
- **Low valence** → Sad / Emotional clusters

These patterns confirmed that the clusters were emotionally meaningful and aligned with human perception.

6.2 Genre Distribution Insights

An analysis of genre distribution within clusters revealed that:

- Clusters were **not genre-pure**
- The same genre often appeared across multiple clusters
- Emotion and genre are **not equivalent concepts**

This observation validates the need for **audio-based emotional modeling** rather than relying solely on genre labels. It also reinforces the importance of unsupervised learning for mood discovery.

7. Recommendation System Design

7.1 Cold-Start Mood-Based Recommendation

For users with no listening history (cold-start users), a mood-driven recommendation strategy was implemented.

◊ **Recommendation Flow:**

1. The user selects a mood (e.g., *party*, *study*, *sleep*)
2. The selected mood is mapped to a corresponding GMM cluster
3. Songs are filtered based on:
    a. Cluster membership
    b. Dominant genres within that cluster
4. Filtered songs are ranked by **popularity**
5. A fallback strategy ensures sufficient recommendations even if the cluster is sparse

This approach ensures that recommendations are:

- Emotionally aligned

- Familiar and trustworthy
- Robust to data sparsity

7.2 Collaborative Filtering (Item-Based)

Since real user interaction data was unavailable, a simulated interaction framework was created.

Key Steps:

- **Generated synthetic user–song interactions**
- **Defined actions: like, listen, skip, listen-like**
- **Assigned numerical scores to reflect engagement intensity**
- **Ensured realistic sparsity patterns**

Outcome:

This enabled:

- **Modeling of preference reinforcement**
- **Simulation of recommendation improvement over time**
- **Integration of mood-based discovery with preference-aware refinement**

7.3 Matrix Factorization using SVD (Collaborative Filtering)

◊ User–Item Interaction Matrix

◊ Created a user–song matrix using simulated interaction scores.
◊ Rows represent users, columns represent songs, values represent engagement strength.
◊ Missing interactions were filled with zero to manage sparsity.

◊ Normalization

◊ Applied user-wise mean normalization to remove rating bias.
◊ Ensured SVD learns relative user preferences instead of absolute scores.

◊ Latent Factor Learning

◊ Applied Singular Value Decomposition (SVD) to extract latent user and song factors.

◊ Learned hidden preference dimensions such as energy level, mood inclination, and danceability.

◊ Prediction Reconstruction

- Reconstructed the matrix to predict interaction scores for unseen user–song pairs.
- Enabled personalized recommendations beyond observed interactions.

◊ Bias-Aware Optimization

- Incorporated global mean, user bias, and song bias.
- Applied SVD on the residual matrix to improve prediction accuracy.

◊ Evaluation

- Evaluated model performance using RMSE on train–test split data.
- Observed reduced RMSE after bias correction.

◊ Hybrid Recommendation Integration

- Combined SVD-based personalized ranking with GMM-based mood clustering.
- Final recommendations balance emotional relevance, personalization, and cold-start robustness.

8. Insights & Business Interpretation

Key Insights

- **Emotional clusters naturally emerge from intrinsic audio features**
- **Mood-based recommendations reduce repetition fatigue**
- **Cold-start users can receive meaningful recommendations without history**
- **Hybrid approaches outperform purely genre-based systems**

Business Impact

- **Reduced song skip rates**
- **Longer listening sessions**
- **Increased playlist creation**

- **Stronger emotional connection with users**
- **Higher user retention and subscription likelihood**

9. Limitations & Future Scope

- ### Limitations

- **No lyrical sentiment analysis**
- **Absence of real user interaction data**
- **Cultural variation in mood perception**
- **Emotional space discretized through clustering**

- ### Future Scope

- **Integrate lyrical sentiment using NLP**
- **Use real streaming interaction logs**
- **Incorporate temporal context (time of day)**
- **Apply reinforcement learning for adaptive personalization**

10. Conclusion

This project demonstrates that advanced machine learning is not about prediction accuracy alone, but about discovering structure, extracting insights, and supporting decision-making in complex, unlabeled environments.

By combining Gaussian Mixture Models with recommendation system principles, we built a hybrid music recommender that understands how music feels, not just what it is called. The system effectively bridges the gap between human emotion and machine understanding, offering a scalable, explainable, and business-relevant solution for modern music platforms.

Declaration

We confirm that:

- This work is original.
- The dataset is self-sourced and approved.
- All project guidelines have been strictly followed.