Presented By **Group 7**

Developing Music Classification Method and Recommendation System

Empowering Music Therapy with Sustainable, Inclusive Practices


Nguyen Mai Anh


Hoang Nguyen Gia Huy


Hoang Nguyen Nhat Phuong

Spotify

Overview:

Spotify is a digital music streaming service founded in 2006 in Stockholm, Sweden, by Daniel Ek and Martin Lorentzon. It offers a vast library of audio content accessible via smartphones, computers, tablets, and gaming consoles, operating on a "freemium" model that provides both free and paid subscriptions.

180
Countries

246M
paying subscribers

626
Monthly Active Users

#1
Streaming Platform

Objective: As Spotify product team, how can we make music therapy accessible via our app?

Market Growth:

US$2.4B
global market valuation

9,1%
CAGR

Key drivers:

Mental Health Challenges
(stress, anxiety, chronic illnesses, etc.)

Digital Health Technologies
(mobile applications, AI/ML-based personalization, VR/AR)

In recent years, alternative therapies like meditation, acupuncture, and creative arts interventions have gained attention for promoting holistic health and well-being (Rebecchini, 2021). Although often seen as a cost-effective complement. However, it faces challenges:

- High costs
- Limited Therapist Availability
- Logistical Barrier
- Misconception

Question Breakdown:

- When can we know if he/she is mentally unstable and needs therapeutic music?
- What can we offer to make the music therapy session frictionless?
- How can we make the integration mindfully of the Marketing and CSR strategy of Spotify?

Objective: As Spotify product team, how can we make music therapy accessible via our app?

The core idea: What if Spotify could not only match your mood - but help you shift it, gently?

*Users often turn to music to manage their emotional states - yet Spotify currently **only mirrors moods**, rather than **guiding emotional transitions**. With mental well-being becoming more central to digital experiences, there's an opportunity to make Spotify a proactive emotional wellness companion, especially during Mental Health Awareness Month.*

Goal

Build a mood-aware music therapy experience using existing session data and clustering models to deliver emotional persona cards and mood-shifting playlists.

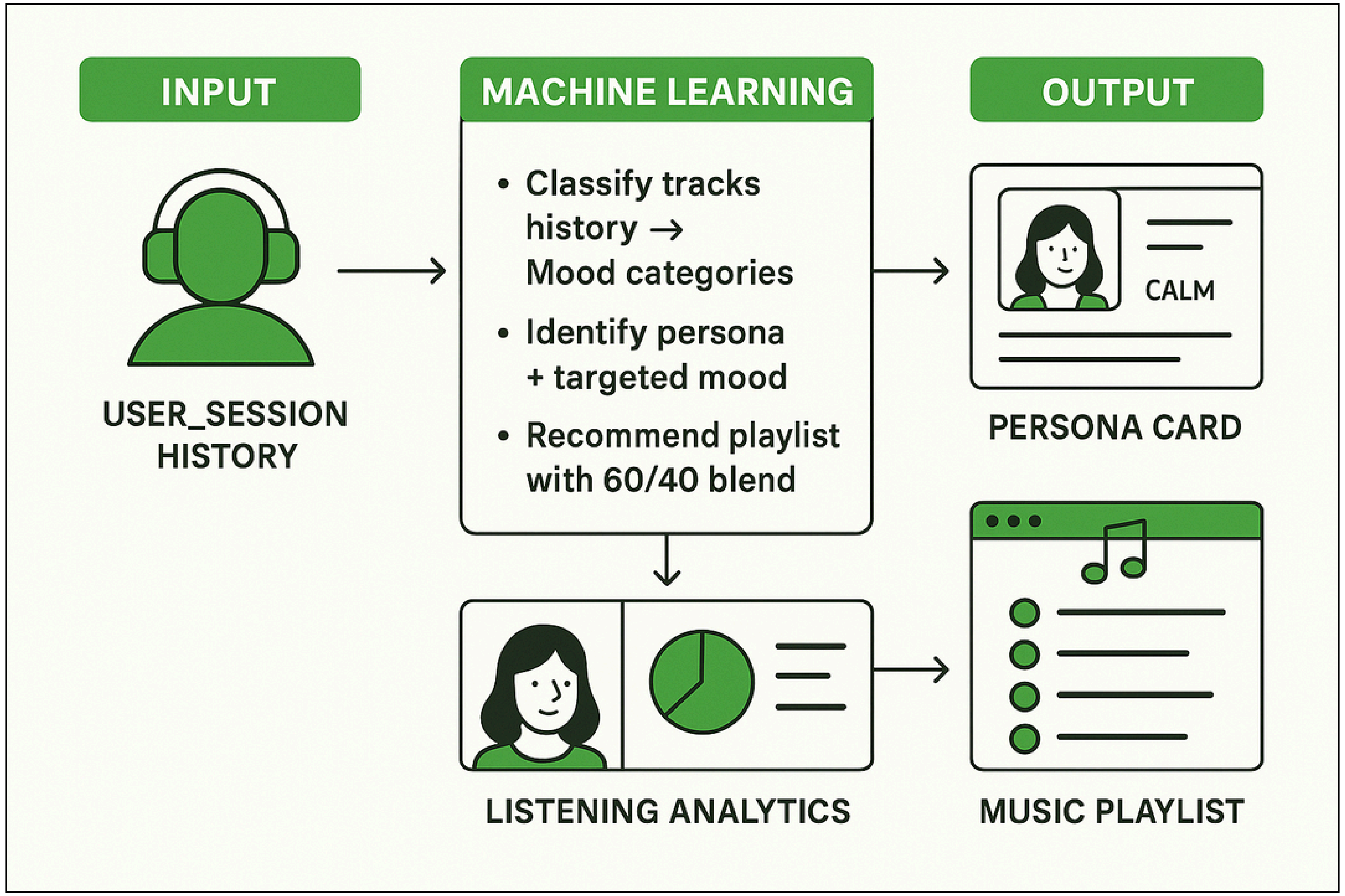
Target users	Why They're Relevant
Gen Z and Millennials	Use music for stress/anxiety coping, often at night, needing reflective mood-based recommendations.
Late-Night Listeners	Seek introspective, emotionally charged sessions, ideal for mood-tailored music.
Mindfulness/ Self-Care Subscribers	Focus on self-care, benefiting from calming, mood-aware playlists that balance familiarity and relaxation.

Objective: As Spotify product team, how can we make music therapy accessible via our app?

Success Metrics

Goal	Metric
Engagement	Increase in playlist completion rate
Emotional Impact	Positive sentiment from reflection prompts ("How did this feel?")
Retention	Uplift in weekly active users in the feature cohort
Personalization Accuracy	Reduction in playlist skip rate post-adjustment
Social/Brand Impact	Social shares or mentions during Mental Health Month

Technical Approach

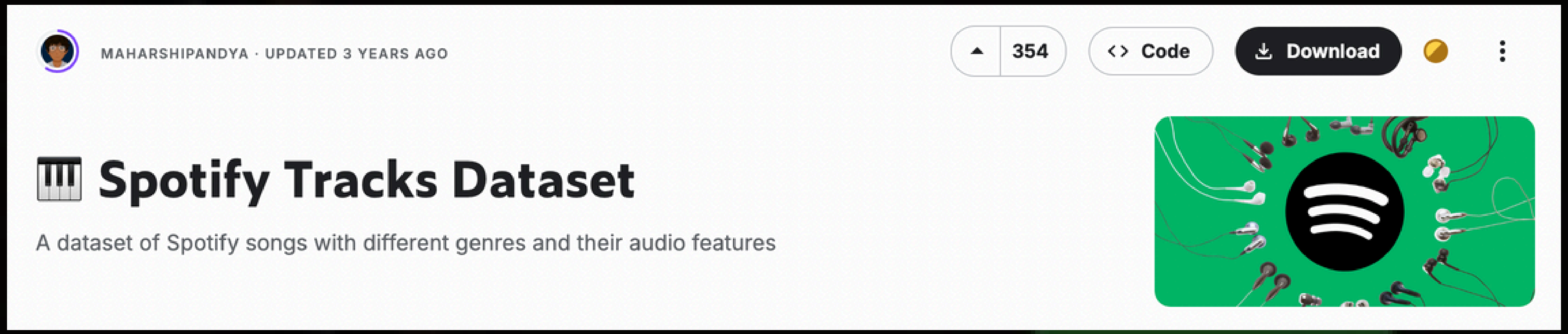


Since there is no available dataset that has specific user sessions with musical features available, we **connect these two tables by using SQL, mapping User_session.Artist, User_session.Track, User_session.Album Name** with the reference data from Spotify

We combine **audio feature data** with **user listening history** to build a therapeutic recommendation system that adapts to individual mood and behavior patterns on Spotify.

1. Spotify Tracks Dataset

Source: kaggle



125 different genres

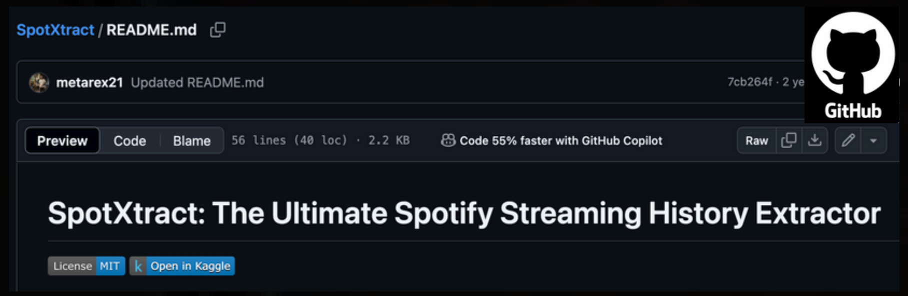
90,000 songs

21 columns

15 musical features

2. SpotXtract – StreamingHistory6

Source



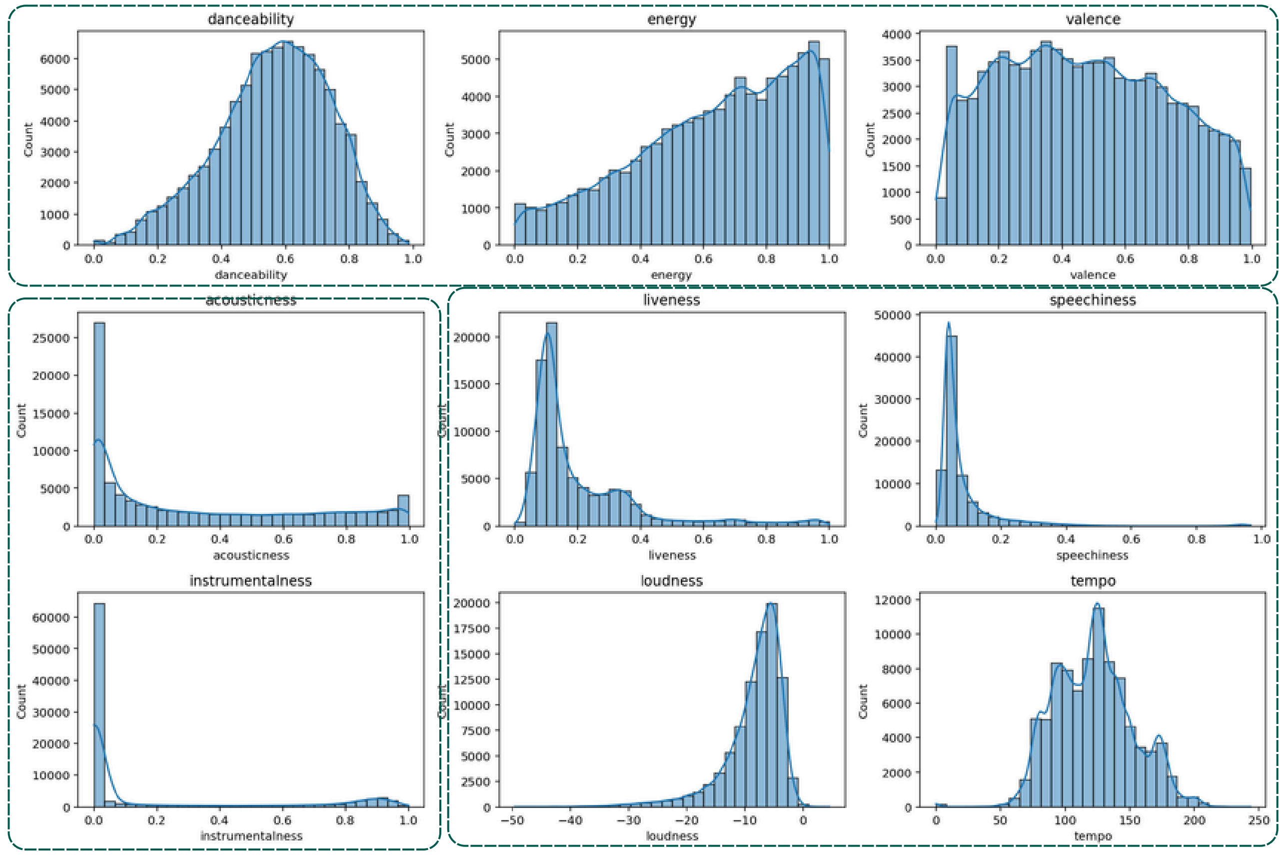
The dataset records listening history of author for the duration of 3 months

Overview

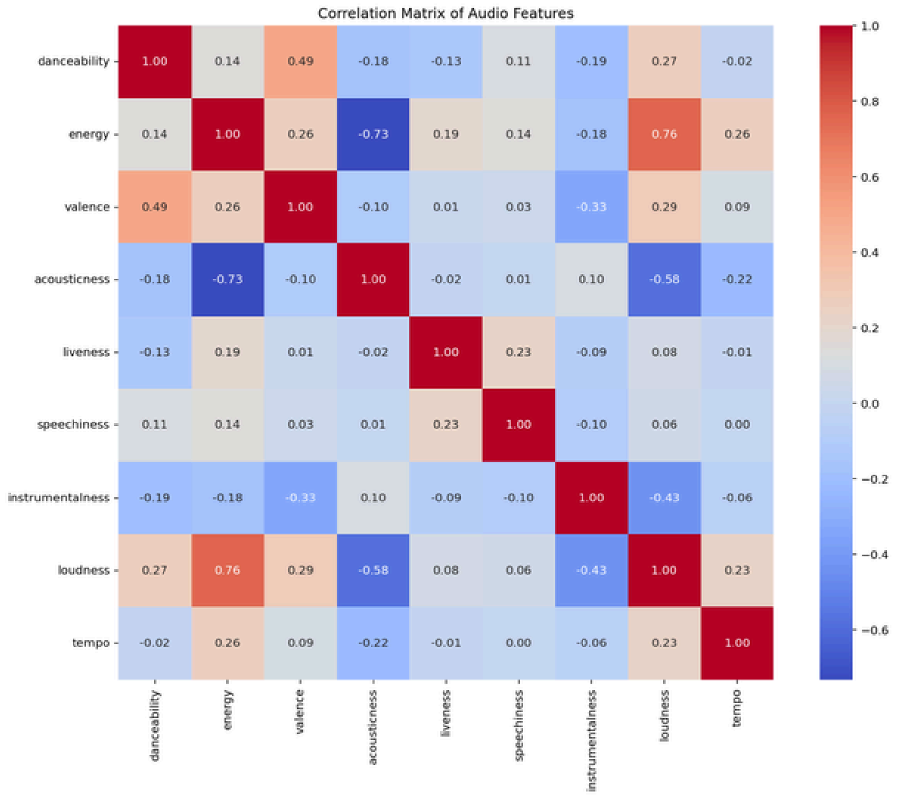
483 unique session ID

4 key information about *(Artist, Track, endTime, msPlayed)*

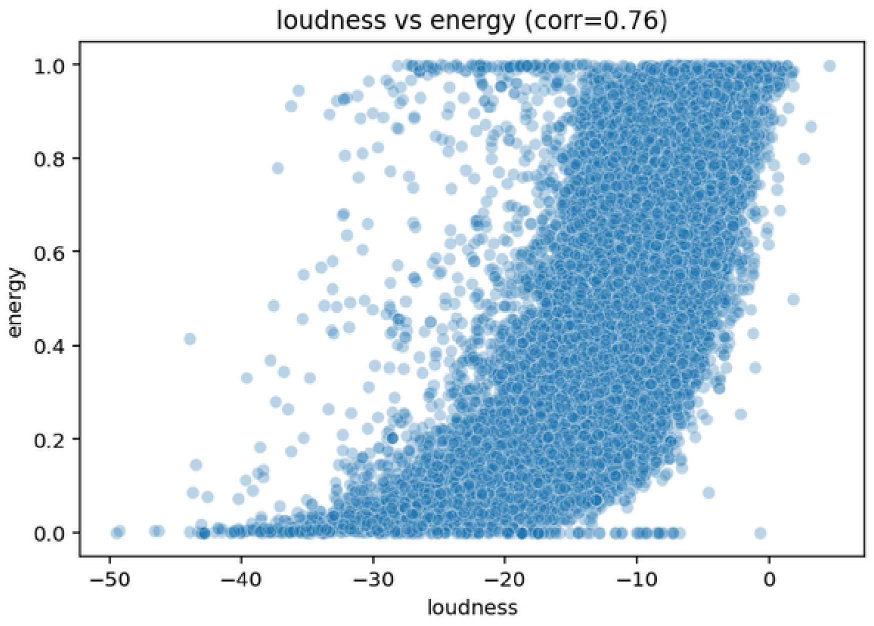
Abstraction/EDA Findings

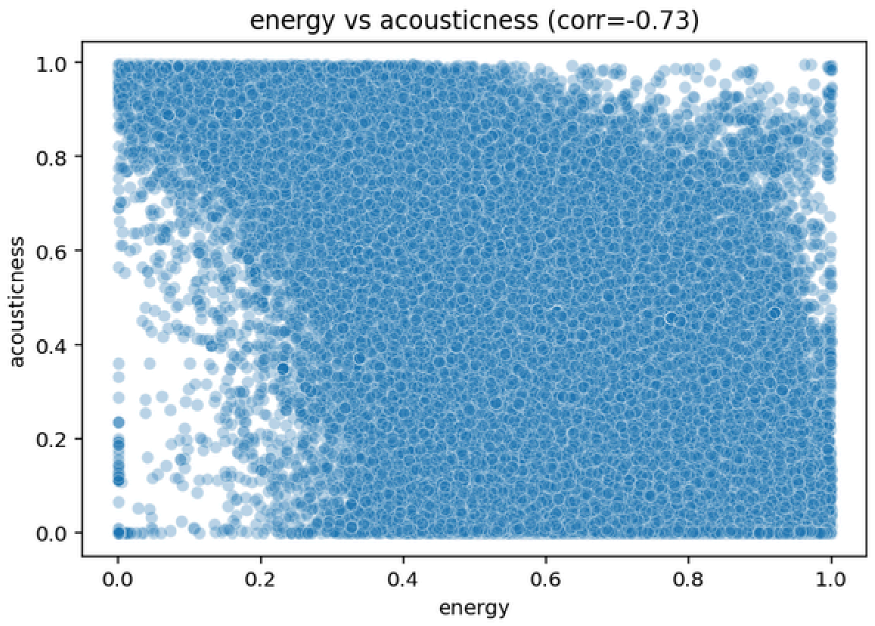


- Tracks are generally **upbeat, energetic, and emotionally positive.**
- Most songs include **vocals and electronic elements** rather than being purely acoustic or instrumental.
- Most tracks having **little spoken content**, but a small subset (e.g., rap or spoken word) showing high speechiness.



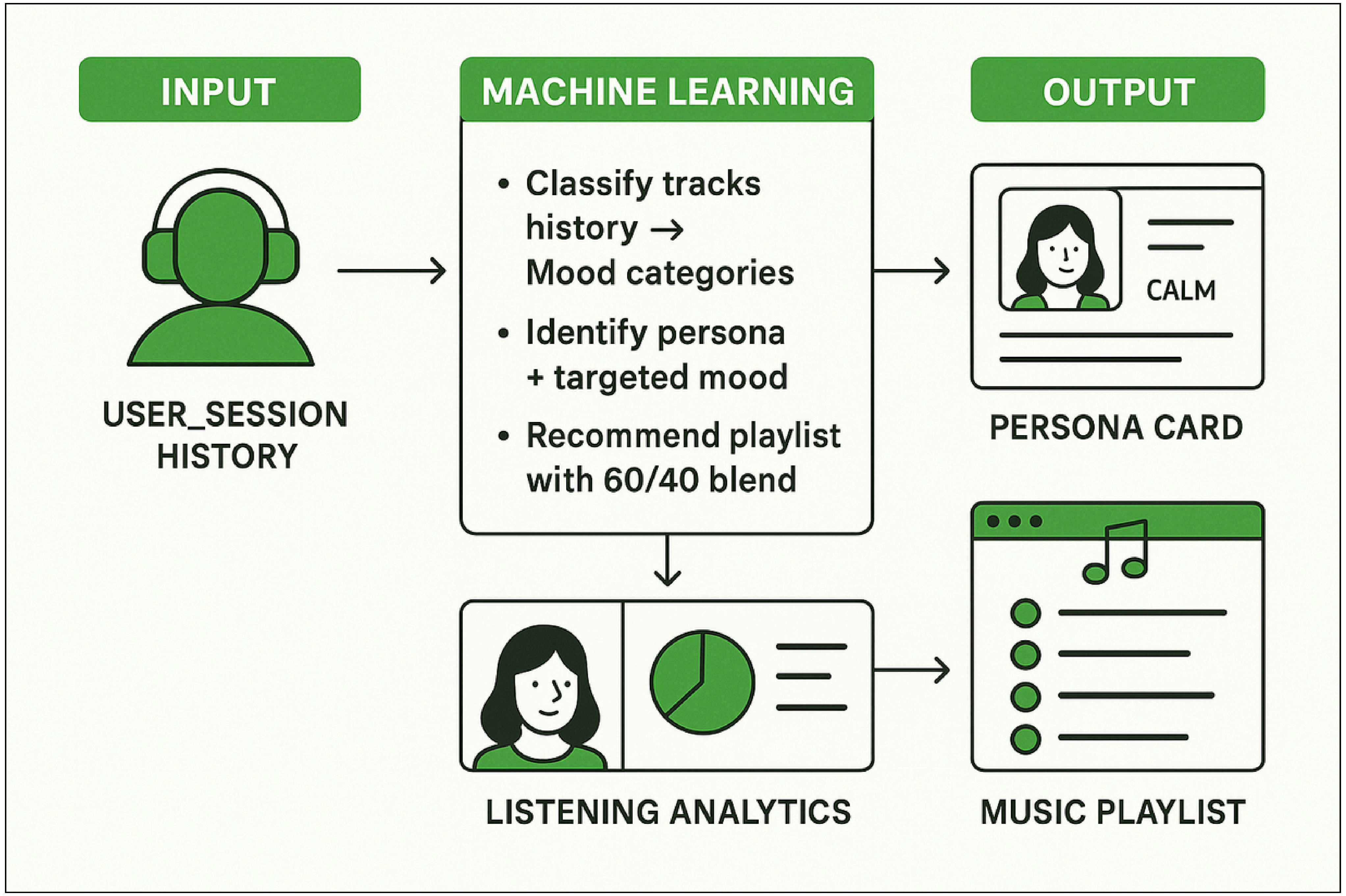
variables with correlation > 0.6





Algorithmic Thinking

Visualized Pipeline



Machine Learning Model

We apply three clustering algorithms (K-means, Gaussian Mixture Model (GMM), and Fuzzy C-means) to classify songs based on mood. We then use the K-Nearest Neighbors (KNN) algorithm to generate recommendations based on user listening sessions.

Clustering Algorithms

- K-Means
- Gaussian Mixture Model (GMM)
- Fuzzy C-Means (FCM)

Recommendation Algorithm

- K-Nearest Neighbors (KNN)

Clustering Algorithms

Purpose

To group songs by emotional similarity using audio features as a basis for mood-based music recommendations and emotional transitions.

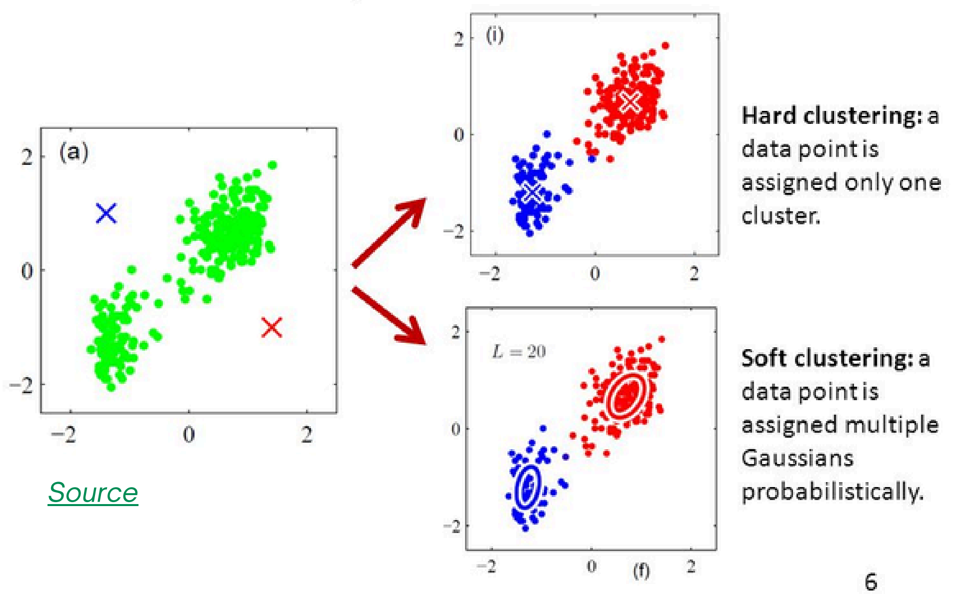
Approach

We used unsupervised learning to group songs based on key features - **energy, acousticness, valance** - which are most indicative of mood classification (Dalida et al.)

Ranking	Audio Features
1	Energy
2	Acousticness
3	Valence
4	Instrumentalness
5	Speechiness
6	Danceability
7	Liveness
8	Mode
9	Loudness
10	Time Signature
11	Key
12	Tempo

Source

Feature	K-Means	GMM	Fuzzy C-Means
Type	Hard clustering	Soft clustering	Soft clustering
How it works	Finds closest center	Shares membership based on distance	Uses probability to assign clusters
Membership	0 or 1	Value between 0 - 1	Probability between 0 - 1
Cluster Shape	Round or equal-sized	Round or equal-sized	Can be oval and stretched



Source

Model Performance

Silhouette Score ?

Measures cohesion and separation of clusters (Rousseeuw, 1987)

Davies-Bouldin Score ?

Evaluates intra-cluster similarity and inter-cluster difference (Davies & Bouldin, 1979)

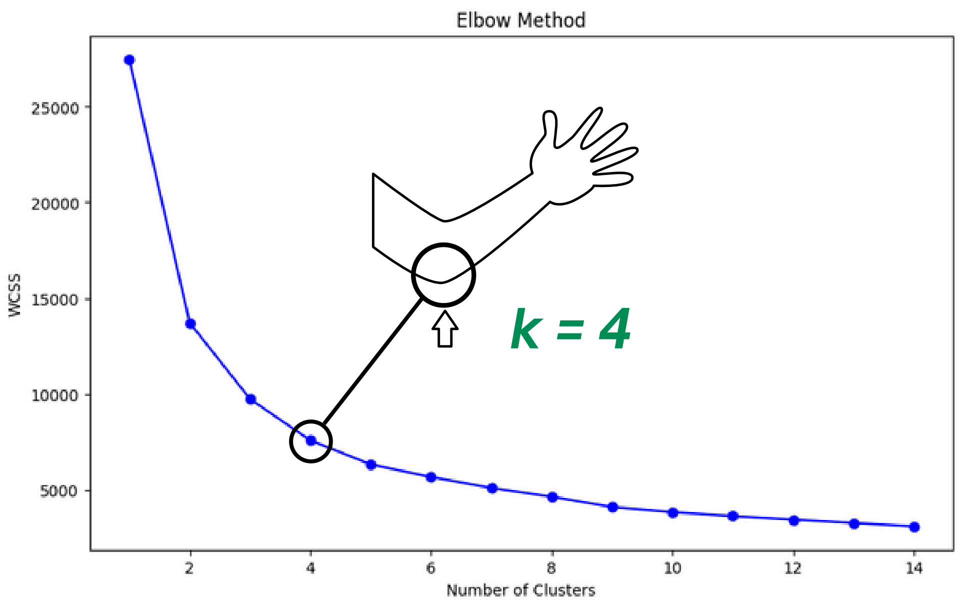
Mood-based Clustering

Purpose

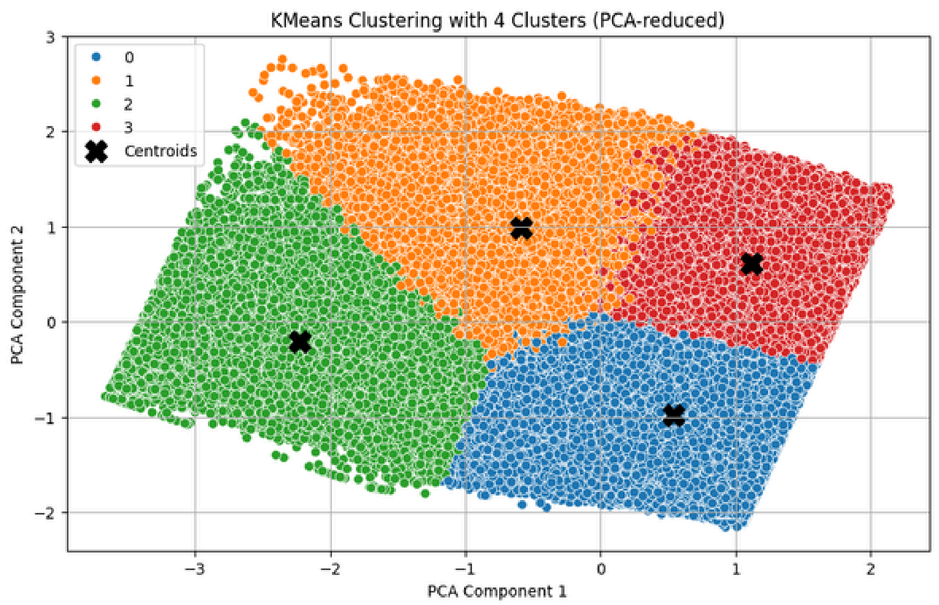
To group songs by emotional similarity using audio features as a basis for mood-based music recommendations and emotional transitions.

Approach

K-means Clustering



*This method helps identify the point (referred to as k) where **adding more clusters no longer leads to significant improvements** in the model.*



***Valence, energy, and acousticness** can group songs by emotional tone fairly well, though some **overlap between moods** is natural.*

```
[ ]  # Show cluster centers for 4 clusters
     print("Cluster Centers (4 Clusters):")
     print(kmeans_4.cluster_centers_)

     Cluster Centers (4 Clusters):
     [[ 0.46974306 -0.68826245 -0.7582459 ]
      [-0.35055471  0.86005016  0.7186202 ]
      [-1.48539521  1.42712478 -0.87520184]
      [ 0.60487687 -0.62739497  0.93681906]]
```

Cluster	Energy	Acousticness	Valence	Mood Summary	Mapped Mood
0	High	Low	Low	Intense, emotional, dark	**Angry**
1	Low-Med	High	High	Relaxed, acoustic, cheerful	**Calm**
2	Very Low	Very High	Very Low	Soft, acoustic, melancholic	**Sad**
3	High	Low	High	Energetic, synthetic, joyful	**Happy**

*Each cluster was labeled with an **emotional state** based on the **mean values of the features** within that cluster.*

Model Performance

Silhouette Score **0.40**

Measures cohesion and separation of clusters (Rousseeuw, 1987)

- Moderate clustering (<0.5)
- Clusters are somewhat distinct, but there's still overlap or ambiguity between them.

Davies-Bouldin Score **0.84**

Evaluates intra-cluster similarity and inter-cluster difference (Davies & Bouldin, 1979)

- A score of 0.84 (<1) is considered good
- Clusters are relatively compact and distinct from each other.

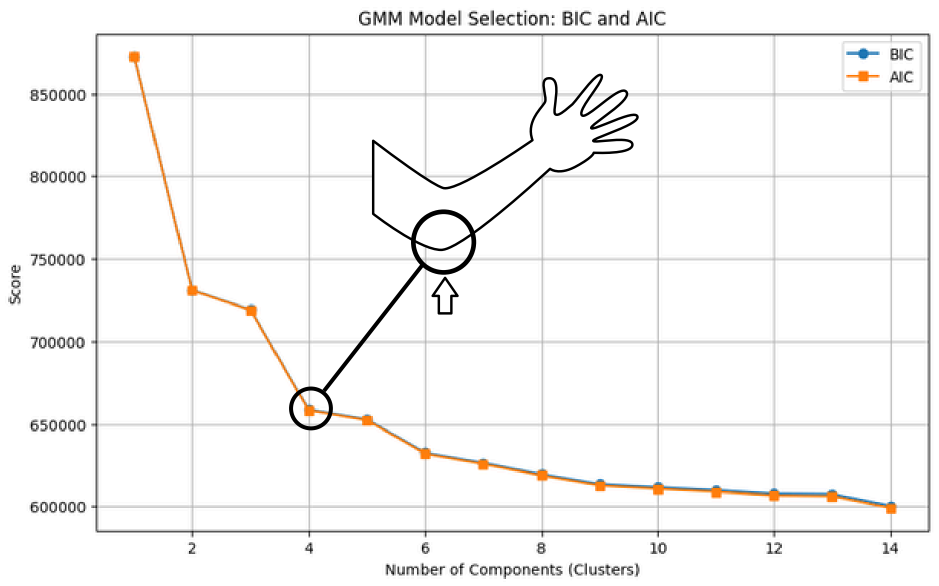
Mood-based Clustering

Purpose

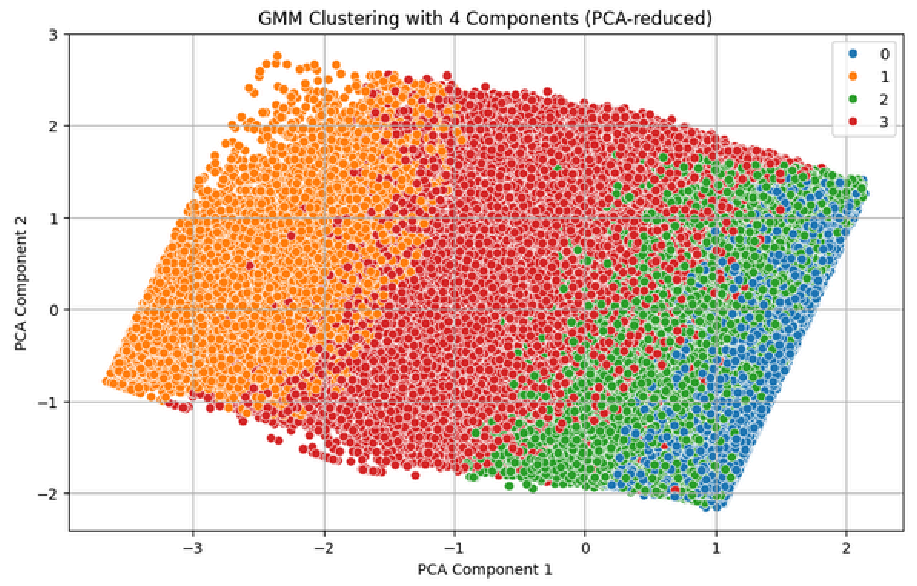
To group songs by emotional similarity using audio features as a basis for mood-based music recommendations and emotional transitions.

Approach

Gaussian Mixture Model (GMM)



We use BIC and AIC to determine ***the optimal number of clusters, which is 4****, where both scores are minimized.*



Valence, energy, and acousticness *perform ineffectively with significant* ***overlap between moods***

```
GMM Cluster Means:
[[ 0.97422087 -0.9415563  -0.26274758]
 [-1.59215991  1.710761   -0.56515562]
 [ 0.42612822 -0.78145525  0.15784037]
 [-0.14560179  0.33573763  0.20873975]]
```

Each cluster center represents a distinct mood profile, summarized as follows:

Cluster	Energy	Acousticness	Valence	Mood Summary	Mapped Mood
0	High	Low	Low	Intense, synthetic, emotional	**Angry**
1	Very Low	Very High	Low	Soft, acoustic, melancholic	**Sad**
2	Moderate	Low	Slightly High	Energetic, upbeat, electronic	**Happy**
3	Low	Moderate	Moderate	Mellow, calm, balanced	**Calm**

Each cluster was labeled with an ***emotional state*** *based on the* ***mean values of the features*** *within that cluster.*

Model Performance

Silhouette Score **0.11**

Measures cohesion and separation of clusters (Rousseeuw, 1987)

- A Silhouette Score of 0.11 means the clustering structure is weak but somewhat present.

Davies-Bouldin Score **2.05**

Evaluates intra-cluster similarity and inter-cluster difference (Davies & Bouldin, 1979)

- A Davies-Bouldin Score of 2.05 indicates moderate to poor clustering quality, with clusters that are not well separated and likely have significant overlap.

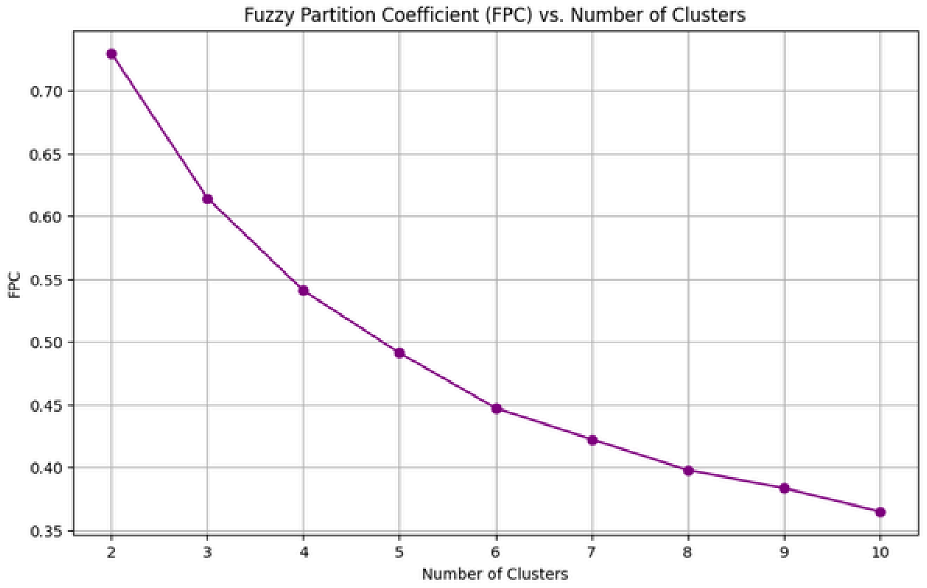
Mood-based Clustering

Purpose

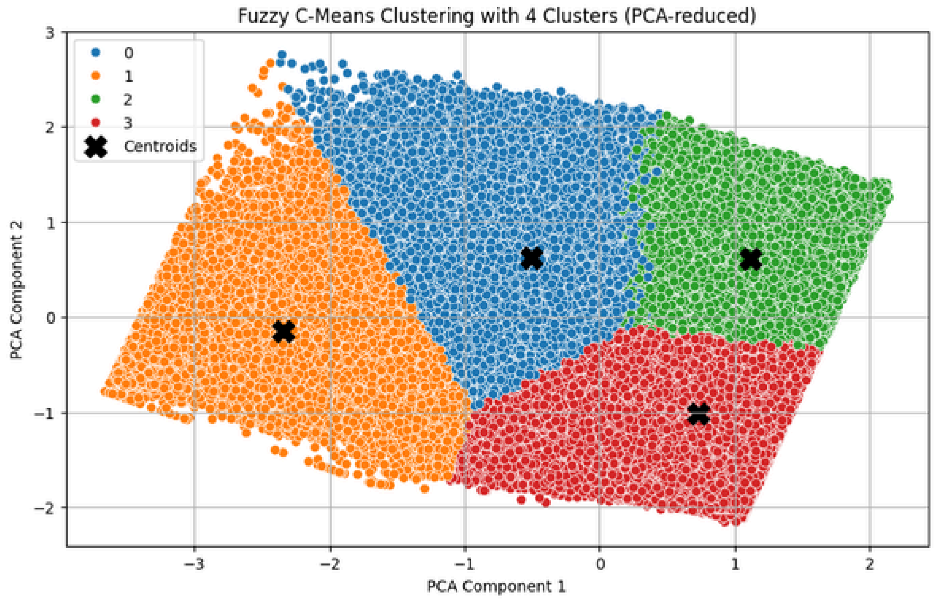
To group songs by emotional similarity using audio features as a basis for mood-based music recommendations and emotional transitions.

Approach

Fuzzy C-Means Clustering



__FPC__ is highest at about 0.73 with __2 clusters__ and drops as the number of clusters grows.



Considering __business goals__, choosing 4 clusters might make more sense even though the FPC is lower.

```
Cluster Centers (4 Clusters, Fuzzy C-Means):
[[-0.3534663   0.5924078   0.42268587]
 [-1.55563782  1.5347167  -0.8594505 ]
 [ 0.63135098 -0.6092562   0.93526093]
 [ 0.65632463 -0.7631558  -0.74425405]]
```

Cluster	Energy	Valence	Danceability	Interpretation	Mapped Mood
0	Low–Moderate	Moderately High	Moderate	Mildly energetic and positive	**Calm**
1	Very Low	Very High	Very Low	Emotionally rich but low energy and rhythm	**Sad**
2	High	Low	Very High	Intense, negative emotion, rhythmic	**Angry**
3	High	Very Low	Very Low	Intense and dark, low rhythmic feel	**Angry**

Each cluster was labeled with an __emotional state__ based on the __mean values of the features__ within that cluster.

Model Performance

Silhouette Score **0.40**

Measures cohesion and separation of clusters (Rousseeuw, 1987)

- This falls in the moderate range (<0.5). It's acceptable, but not highly separated.

Davies-Bouldin Score **0.86**

Evaluates intra-cluster similarity and inter-cluster difference (Davies & Bouldin, 1979)

- A Davies-Bouldin Score of 0.86 (<1) indicates fairly good cluster separation and compactness.

Clustering Result Comparision

Model	Silhouette Score	Davies-Bouldin Score	Cluster Type	Pros	Cons
KMeans	**0.40**	**0.84**	Hard	- Simple, fast, and interpretable - Well-separated, compact clusters	- Assumes spherical clusters - Hard assignments may ignore ambiguity
GMM	**0.11**	**2.05**	Soft (probabilistic)	- Captures cluster uncertainty - Flexible shapes	- Poor separation - High overlap - Sensitive to initialization
Fuzzy C-Means	**0.40**	**0.86**	Soft (fuzzy)	- Models emotional ambiguity - Comparable structure to K-Means	- Slower convergence - Needs fuzzy parameters tuning

K-Means *delivered the best-defined clusters and* ***Fuzzy C-Means*** *offered comparable structure with the added benefit of modeling emotional ambiguity, while* ***GMM*** *performs poorly in this context.*

Decision Trees

Context Given the mood cluster from the last model, we want to identify one's need for therapeutic music via their listening behaviors (mood, frequency) (Yoon, 2020)

Objective Developing a viral tactic to engage Spotify users into joining Mental Awareness Month

Strategy 06 Personalized personas card to describe user's mental state, reflected by their listening session in Spotify. These personas are designed based on Carl Jung's Theory of Personality & Spotify Wrapped 2022.

*names are subjected to change

Which mood catgory did user listen to most frequently?

Sad | Happy | Angry | Calm

Did one category dominate (>50%) the listener's mood?

One category with >= 70%	Main category with 40-60%, the rest are distributed among all three	None category are over 40%
Replayer	Main category are	Voyager
Secret cards	Sad: Blue Horizon Happy: Fan Club Angry: Fire Storm Calm: Chill Guy	Secret cards

Content-based KNN recommendation

Context Identified the persona of listeners, we move on to recommending therapeutic playlist that act as mood stabilizer

Objective Developing a frictionless therapeutic playlist for users.

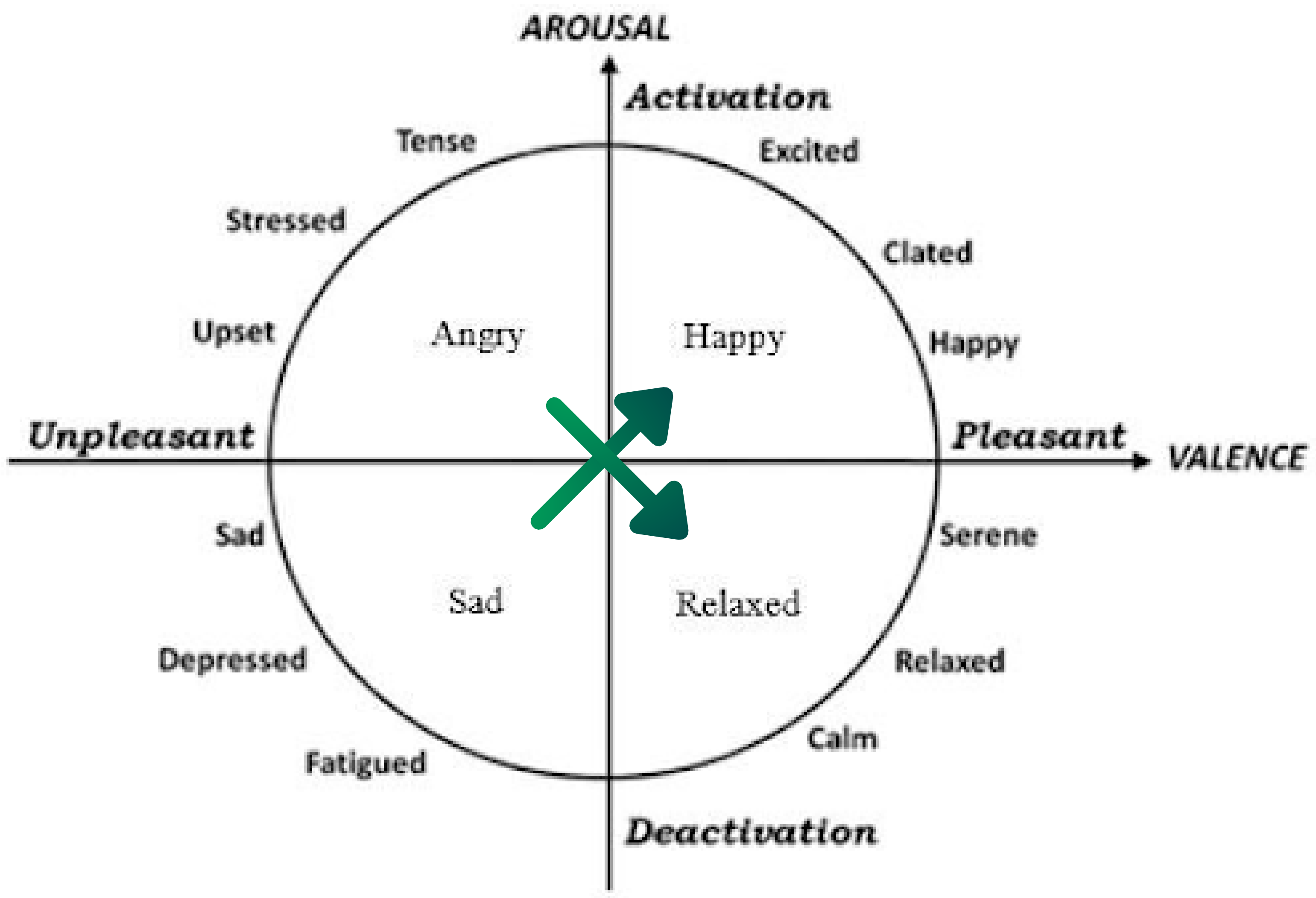


Figure derived from Russel Model of Affect and British Government Health Department Publication

Strategy

A 60-minute playlists by combining songs from the user's dominant mood cluster with tracks from a targeted mood cluster to support emotional regulation, following the Russell Model of Affect.

Components

Target moods | **Current moods**

Sample output

=========Your recommended playlist=======

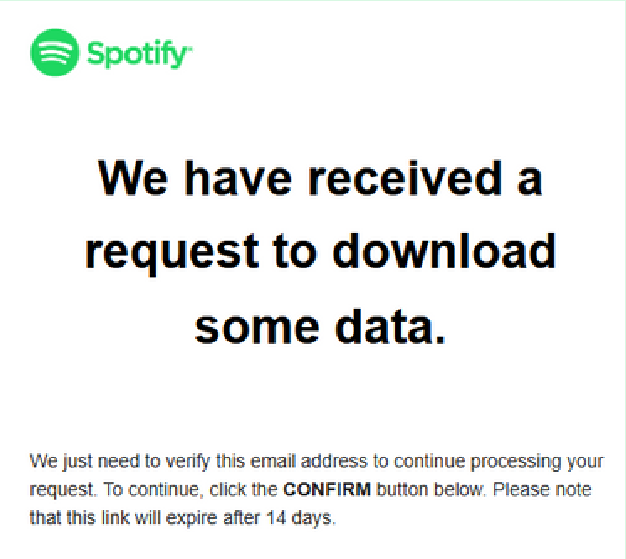
	trackName	artistName	energy	acousticness	valence
72	oops	clide	0.678	0.139	0.702
7	beautiful girls	sean kingston	0.661	0.150	0.769
9	black hole	griff	0.640	0.142	0.658
33	do do do	dansu	0.644	0.137	0.801
14	first night	elijah woods	0.584	0.194	0.727

n = 5, students from all three colleges

H0: The personalized playlist doesn't boost user mental experience

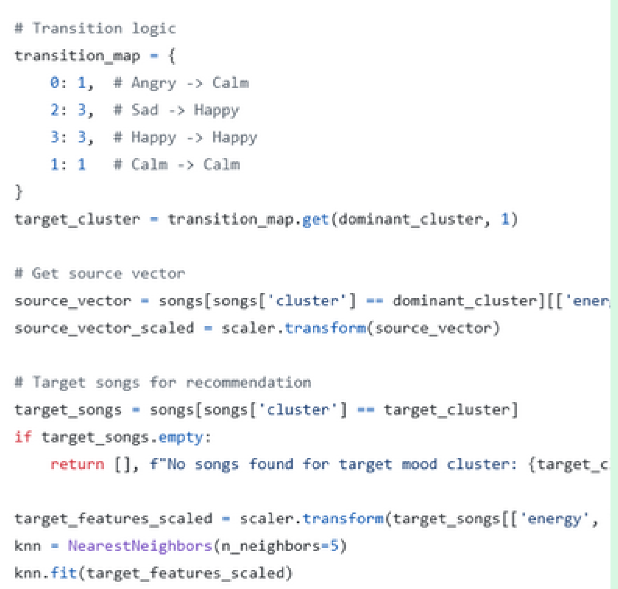
Qualitative

Export interviewee listening history ▶ Run through the model ▶ Generate a common playlist ▶ Record feedbacks


Spotify

We have received a request to download some data.

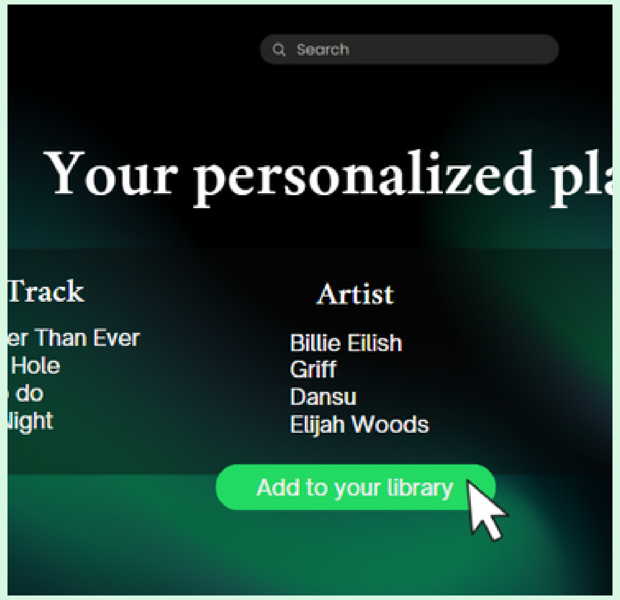
We just need to verify this email address to continue processing your request. To continue, click the **CONFIRM** button below. Please note that this link will expire after 14 days.


```
# Transition logic
transition_map = {
    0: 1,  # Angry -> Calm
    2: 3,  # Sad -> Happy
    3: 3,  # Happy -> Happy
    1: 1   # Calm -> Calm
}
target_cluster = transition_map.get(dominant_cluster, 1)

# Get source vector
source_vector = songs[songs['cluster'] == dominant_cluster][['ener
source_vector_scaled = scaler.transform(source_vector)

# Target songs for recommendation
target_songs = songs[songs['cluster'] == target_cluster]
if target_songs.empty:
    return [], f"No songs found for target mood cluster: {target_c

target_features_scaled = scaler.transform(target_songs[['energy',
knn = NearestNeighbors(n_neighbors=5)
knn.fit(target_features_scaled)
```



**Disclosed conversation due to data privacy*

Result

4/5 participants enjoy the customized therapeutic experience

4/5 agree on the social impact of this campaigns

100% participants would share their persona stats *n = 41, students from all three colleges from Arts Appreciation Class*

H0: The personalized playlist doesn't boost user mental experience

Quantative

Take the dominant mood from survey

Identified target mood

Good normal
nonchalant happi

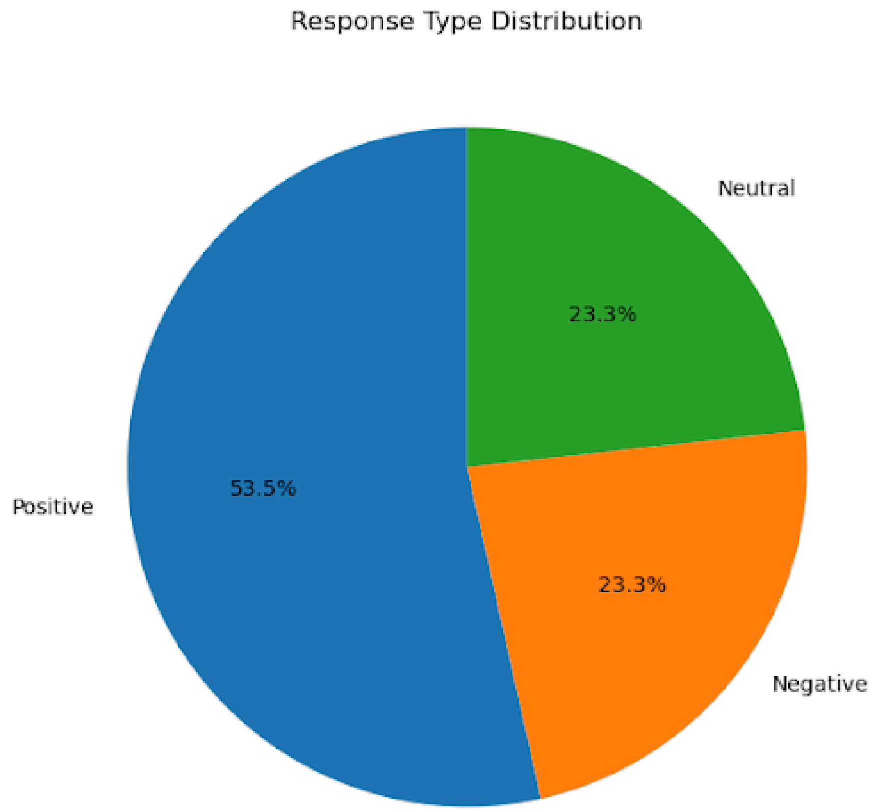
Happy

Pick randomly from happy mood cluster

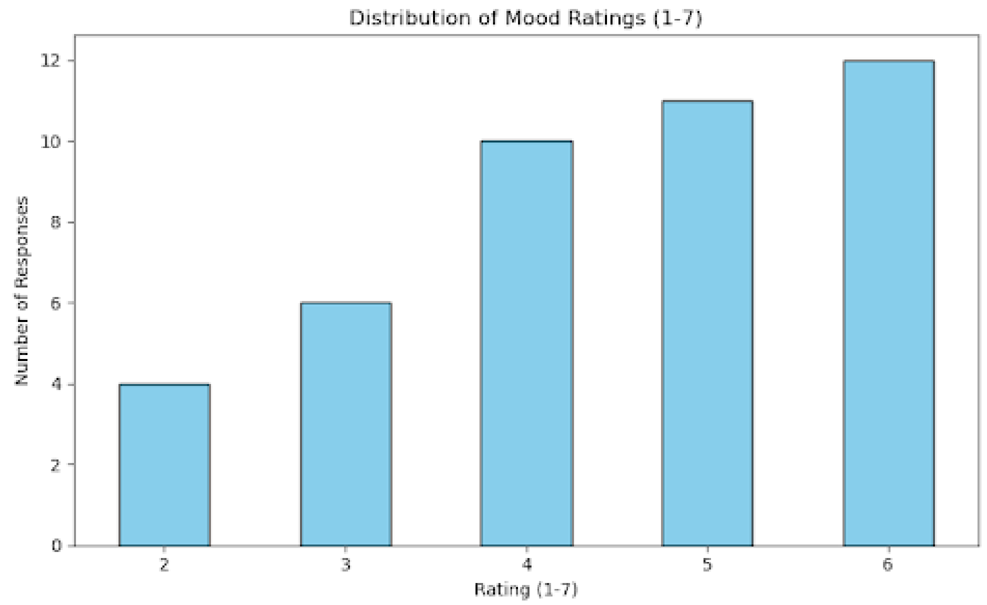
Play and record feedbacks

College of Arts and Sciences	–	Sad	4 / 7
College of Business Management	Art therapy, art expression, art cure methods	Sad	6 / 7

Result



While over-half enjoy the session, there are 23.3% of participants dislike the experience



Reject the **null hypothesis**, personalized playlist works better

User Flow

Input

Passive Listening

Example:

User A listens to 3 songs:

- Song A, B → Cluster 0 (Angry)
- Song C → Cluster 2 (Sad)

Note*: The system will tracks user listening history over 7 days.*

Persona Detection

Following the decision tree to classified user as one of the persona:

- 66% Angry, 33% Sad → classified as **“Angry”**

Output

Persona Card Generation

Element	Example Output
Name	“The Firestorm”
Mood Pie Chart	66% Angry, 33% Sad
Tagline	“You’ve been riding intense and deep waves this week.”
Prompt	“Try something calmer?”

Personalized Playlist Recommendation

Therapeutic Blend:

- 60% songs from user’s recent mood clusters (0 = Angry, 2 = Sad)
- 40% from target cluster (e.g., Cluster 1 = Calm)

Playlist Name Examples: “The Chill Storm”, “Cool Down Without Giving Up the Vibe”

Optional: let users slide the ratio (60/40 to 70/30 or 50/50)

Feedback Prompt

“How did this mix make you feel today?”

Also collect the implicit feedback loop: which transitions they accept or skip

Search

Hi Professor Kok,

Are you ready to explore yourself ?

Next

Limitation & Further improvement

Lack of Generalization

The study was conducted within the scope of VinUniversity, leading to lack of application to other context

Evaluating Outcomes

The outcome was subjectively measured by self-reported survey. External factors may greatly influence their behavior

Subjective Biases

The model was approached from a non-professional viewpoint, leading to assumptions of mental health

Future Work

Involving domain experts to ensure therapeutic credibility

Designing clear focus groups to better personalise the product

Supplement self-report measures with behavioral metrics

Incorporate a hypothesis-driven feedback loop

VINUNIVERSITY

THANK YOU !