

MOODIFY

ENGLISH SONGS MOOD CLASSIFIER

Akshit V Bansal, Navya Nandini Dhody, Saanvi Bhaskar, Sia Jain





Spotify users create over 4 billion playlists, and the majority of them are mood-based: titles like 'Happy Vibes' and 'Sad Nights' outperform many artist-based playlists.



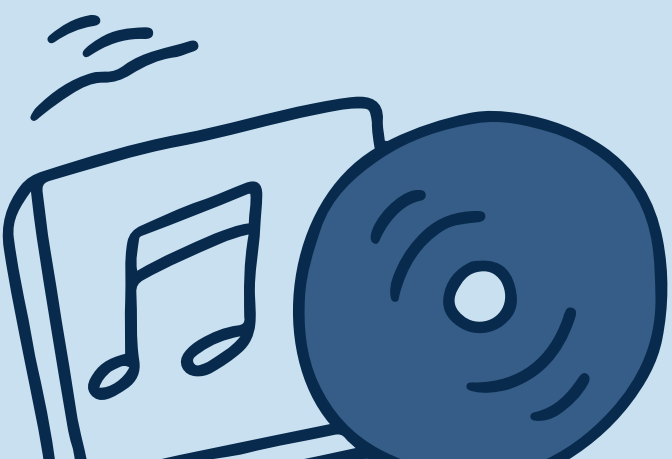
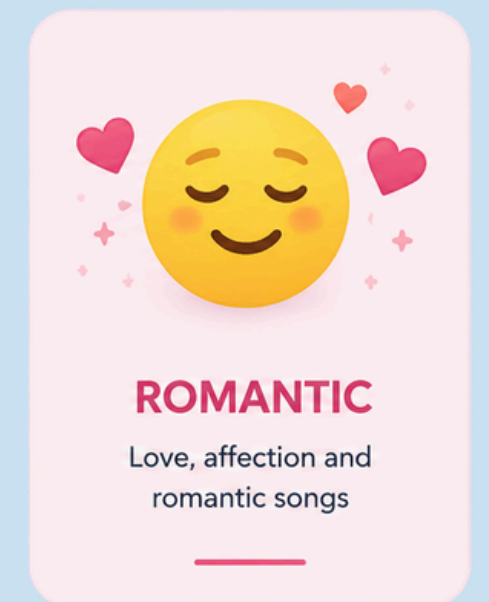
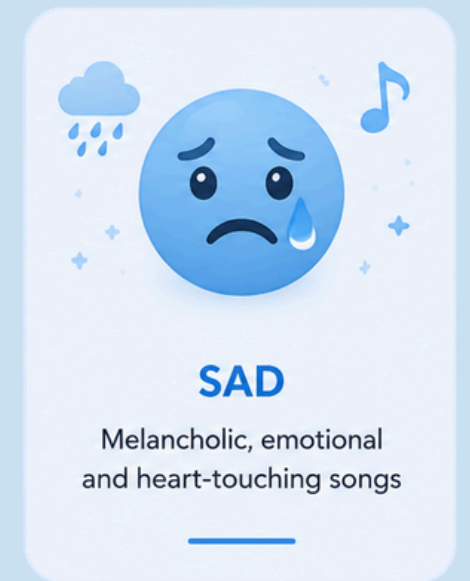
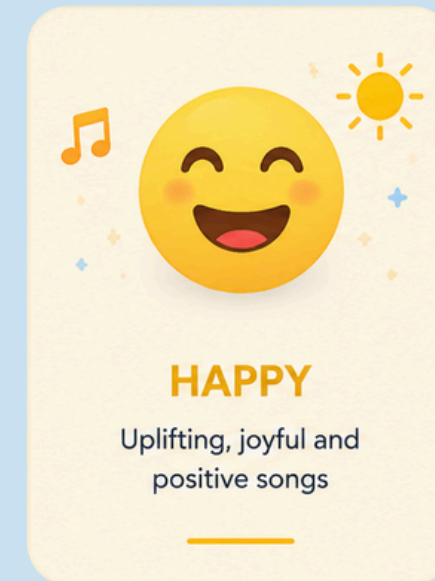
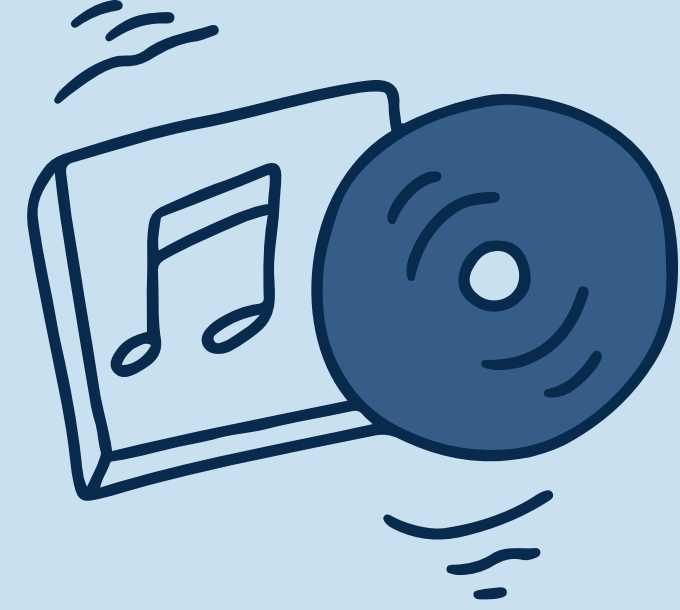
**PROBLEM
STATEMENT**

PROBLEM BEING SOLVED

Music recommendation systems often rely on genre-based classification, which fails to capture the emotional essence of songs, especially in Hindi and Hinglish music where multiple moods and narratives overlap. Existing systems struggle to identify nuanced emotions directly from audio signals.

Our project aims to:

- Classify songs into four moods: Happy, Sad, Excited, and Romantic
- Combine physical audio features with emotional representations
- Improve personalization in recommendation and voice-assistant systems
- Enhance representation of Indian music in AI-driven platforms



POTENTIAL IMPACT

TECHNOLOGICAL IMPACT

- Advances scalable mood classification systems
- Encourages reproducible Indian-focused MIR research

USER IMPACT

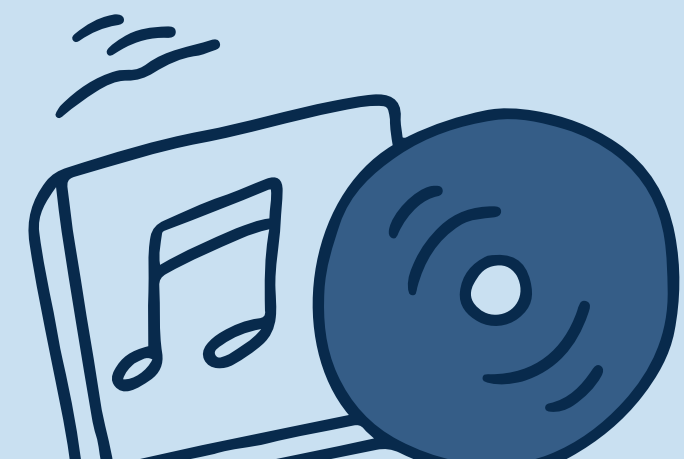
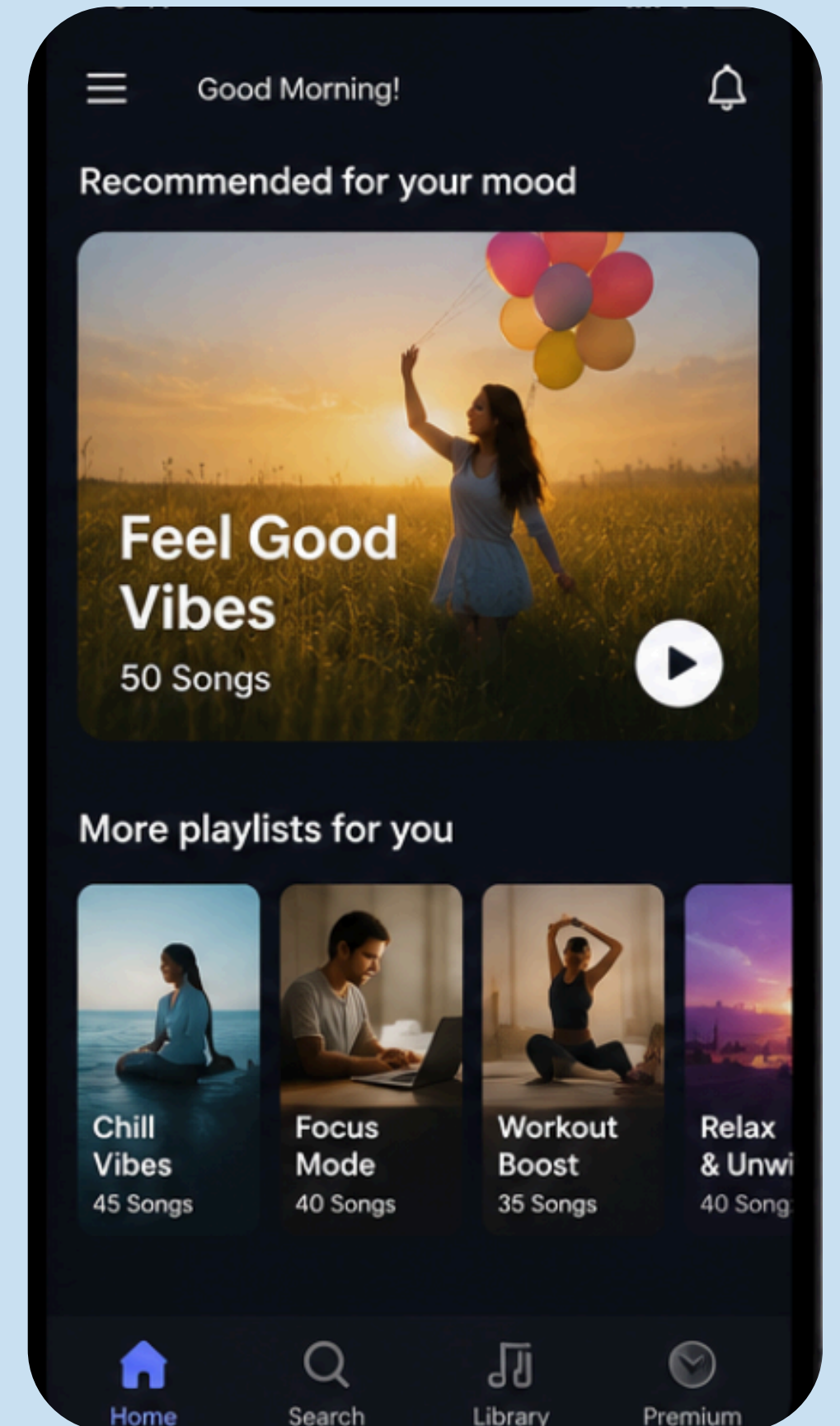
- Enhances mood-based music discovery
- Improves listening experience and engagement

INDUSTRIAL IMPACT

- Enables automated music tagging and recommendation
- Helps streaming platforms improve personalization

POTENTIAL APPLICATIONS

- Mood-based music recommendation systems
- Smart assistants & voice interfaces
- Indian music streaming platforms
- Mental health & wellness applications
- Media and content production
- Music Information Retrieval (MIR) research



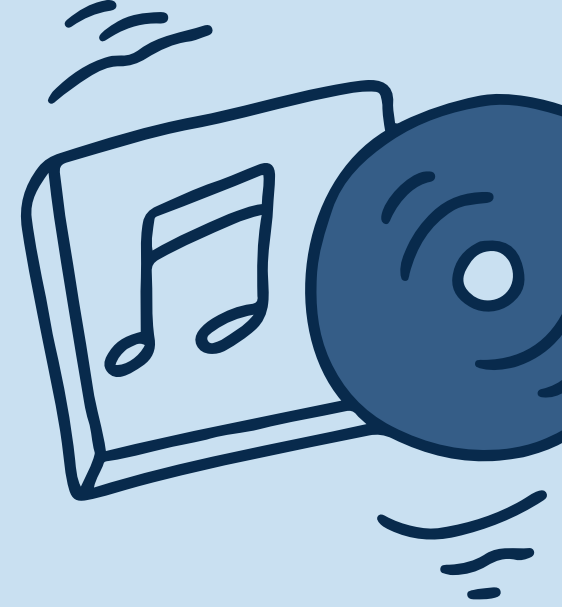
A conceptual image for a literature review. It features a magnifying glass with a silver frame and a grey handle, positioned over a document. The document contains several blue bar charts and line graphs with data points. A silver pen lies diagonally across the right side of the document. The background is a light blue color with a subtle grid pattern. The text 'LITERATURE REVIEW' is centered in a bold, dark blue font.

LITERATURE REVIEW

PAPER
1

**LYRICS-BASED MUSIC GENRE CLASSIFICATION USING A
HIERARCHICAL ATTENTION NETWORK**

Alexandros Tsaptsinos
ICME, Stanford University, USA



**WHAT OTHER RESEARCHES
DID TO SOLVE THE PROBLEM?**

- Proposed a Hierarchical Attention Network (HAN) for lyric-based music classification.
- Analysed lyrics at word, line, and song level using attention mechanisms.
- Compared performance with Logistic Regression and LSTM models.

**DEVELOPED SOLUTIONS THAT
CAME OUT OF THOSE STUDIES**

- HAN achieved higher classification accuracy on large multi-genre datasets.
- Attention layers improved interpretability by identifying important lyrical patterns.
- Best accuracy reported: 46.42% on the 117-genre dataset.

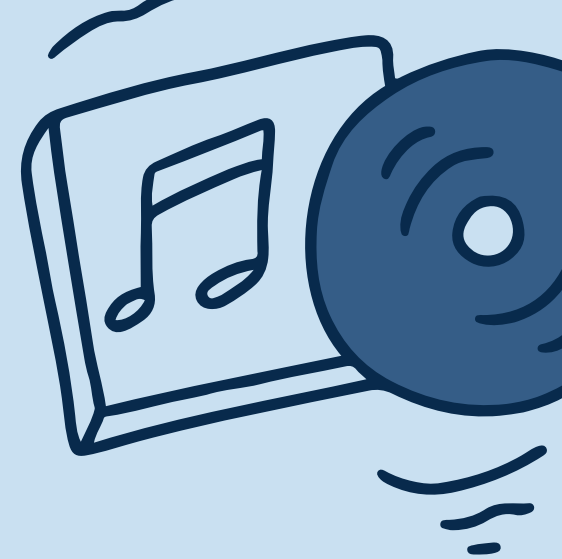
PAPER 2

Enhancing Accuracy and Performance in Music Mood Classification through Fine-Tuned Machine Learning Methods

Shital Shankar Gujar¹, Dr. Ali Yawar Reha²

¹Computer Engineering, Pacific University, Udaipur, sgujar03@gmail.com

²Management, Asst. Prof, Paher University, Udaipur, Ay_reha@yahoo.com



WHAT OTHER RESEARCHES DID TO SOLVE THE PROBLEM?

- Used Music Information Retrieval (MIR) techniques for automated mood classification.
- Extracted audio features such as MFCC, chroma, spectral, and tempo descriptors.
- Applied ML models including SVM, KNN, and ensemble classifiers for emotion recognition.

DEVELOPED SOLUTIONS THAT CAME OUT OF THOSE STUDIES

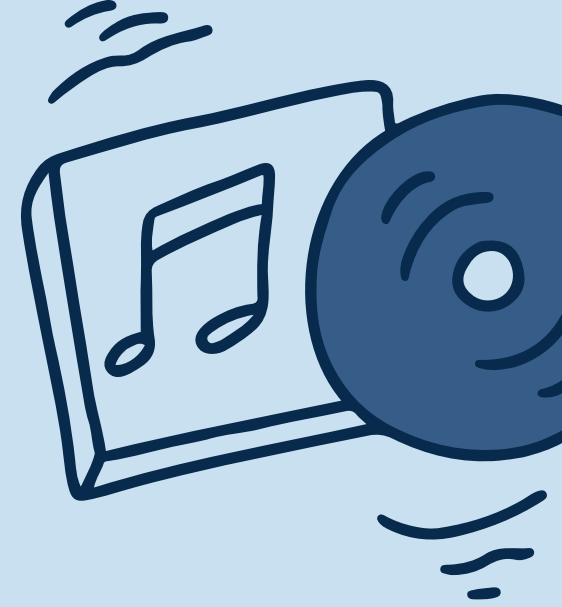
- Combined acoustic analysis with the Valence–Arousal (VA) emotional model.
- Mapped songs into a continuous emotional space instead of fixed labels.
- Improved recommendation and personalization using feature engineering and ensemble learning.

Joint sentiment analysis of lyrics and audio in music

Lea Schaab¹, Anna Kruspe²

¹ *Technische Hochschule Nürnberg Georg Simon Ohm, Email: schaable79981@th-nuernberg.de*

² *Munich University of Applied Sciences, Email: anna.kruspe@hm.edu*



WHAT OTHER RESEARCHES DID TO SOLVE THE PROBLEM?

- Researchers explored music emotion recognition (MER) using both audio and lyrical sentiment analysis.
- Multiple text models and an audio-based CNN model were combined using multimodal fusion techniques.
- Emotions were mapped using the valence–arousal model for sentiment classification.

DEVELOPED SOLUTIONS THAT CAME OUT OF THOSE STUDIES

- Evaluated separate audio-only, text-only, and fusion-based sentiment models.
- Best performance was achieved using a 60% audio + 40% text weighted fusion strategy.
- Results showed that combining lyrics and audio improved emotion recognition compared to using a single modality.

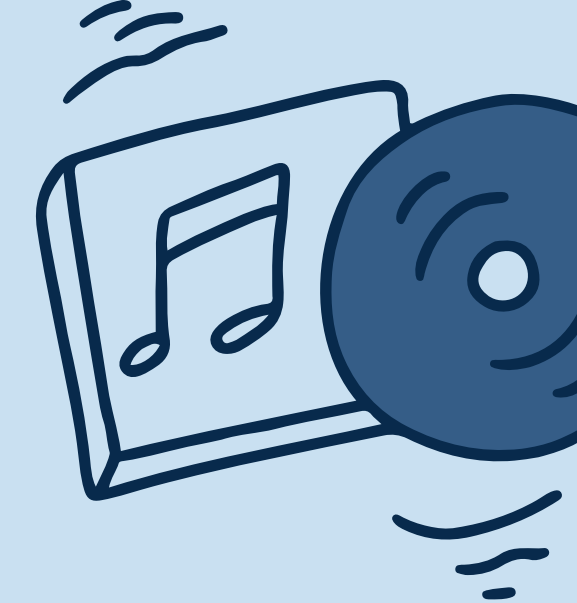
PAPER

4

Music Emotion Maps in Arousal-Valence Space

Jacek Grekow

Faculty of Computer Science, Bialystok University of Technology,
Wiejska 45A, Bialystok 15-351, Poland
j.grekow@pb.edu.pl



WHAT OTHER RESEARCHES DID TO SOLVE THE PROBLEM?

- Modeled music emotion recognition as a regression problem using the Valence-Arousal (VA) emotional model
- Extracted features such as MFCCs, spectral centroid, spectral flux, spectral rolloff, and chroma features using Essentia and Marsyas
- Applied regression models including SMOreg (SVM Regression), REPTree, and M5P for predicting emotional scores

DEVELOPED SOLUTIONS THAT CAME OUT OF THOSE STUDIES

- Created emotion maps to track changes in valence and arousal over time
- Used sliding window segmentation for dynamic emotion tracking
- Improved emotion prediction using feature selection and continuous affective modeling instead of rigid categorical labels

Research Paper	Key Limitations
<p>Joint Sentiment Analysis of Lyrics and Audio in Music (Schaab & Kruspe, 2024)</p>	<ul style="list-style-type: none"> • Small dataset size • Emotional labels were subjective • Audio and lyrics sometimes conveyed different emotions
<p>Lyrics-Based Music Genre Classification Using HAN (Tsaptsinos, 2017)</p>	<ul style="list-style-type: none"> • Focused only on lyrics • Ignored audio/emotional features • Genre classification became difficult with many classes
<p>Music Emotion Maps in Arousal-Valence Space (Grekow, 2015)</p>	<ul style="list-style-type: none"> • Small manually labelled dataset • Emotion prediction depended on human annotation • Valence prediction was less accurate than arousal
<p>Enhancing Accuracy and Performance in Music Mood Classification (Gujar & Reha, 2024)</p>	<ul style="list-style-type: none"> • Relied heavily on handcrafted features • Limited focus on multilingual music • Scalability and real-world deployment were not explored

0x73746172, 0x61286469, 0x6e742822, 0x6f727922, 0x09232073, 0x20726f79, 0x79203d20, 0x28293a0a, 0x7479202b, 0x205b746f, 0x6f74616c, 0x69206e65, 0x69746572, 0x63657620, 0x3a762066, 0x5b305d20, 0x656b207a, 0x0c052864, 0x74737769, 0x72656374, 0x27257327, 0x20252064, 0x65737461, 0x616c7479, 0x302320, 0x0909746f, 0x3d20765b, 0x74616c20, 0x20646f73, 0x6972616a, 0x0a096966, 0x6f72206b, 0x3e3d206c, 0x61207372, 0x09746f74, 0x69726563, 0x74682822, 0x6f72792c, 0x20697320, 0x69726563, 0x76692070, 0x0a09746f, 0x325d0a09, 0x20225673, 0x6567616a, 0x20707265, 0x206c696d, 0x2c762069, 0x696d6974, 0x65647374, 0x616c203d, 0x69726563, 0x746f7279, 0x6f767a65, 0x74616c20, 0x2b5d2076, 0x73616c65, 0x65682070, 0x6f206c69, 0x6b6f206b, 0x6974203e, 0x6e207361, 0x7d0a0909, 0x76612070, 0x20300a09, 0x20762069, 0x616c6573, 0x290a0965, 0x2920616e, 0x293a0a09, 0x6573290a, 0x61206669, 0x290a0909, 0x74656b20, 0x3d20300a, 0x7620696e, 0x5b305d0a, 0x735b225f, 0x726f6461, 0x6f647374, 0x6d697461, 0x6f70696a, 0x20313a0a, 0x6c65732e, 0x2320697a, 0x726f6461, 0x09746f74, 0x6e207361, 0x225f5f54,

DATASET CURATION & FEATURES PREPROCESSING

DATASET COLLECTION

DATASET DESCRIPTION

- Manually curated dataset of Hindi and Hinglish songs
- Built to study mood classification in Indian popular music
- Each song is labeled based on four dominant moods:
Happy, Sad, Romantic, and Excited

DATASET SIZE

Metric	Value
Total Songs	~1600+
Mood Classes	4
Audio Format	MP3
Sampling Rate	22,050 Hz
Average Song Length	~3–4 minutes
No. of Features Extracted	50+

DATA COLLECTION CONSIDERATIONS

MOOD CLARITY

- Songs with clear dominant moods were selected.
- Mood labels were generated using 3 LLMs and manually verified.
- This reduced bias and subjectivity in labeling.

METADATA

- Songs collected across:
- decades (1950–2020s)
 - different artists
 - different musical styles

AUDIO QUALITY

Only clean MP3 recordings were used to avoid noise affecting feature extraction.

Standardized audio format used for preprocessing consistency.

ETHICAL CONSIDERATIONS



RESEARCH & EDUCATIONAL USE ONLY

The dataset has been collected and used strictly for research and educational purposes.



COPYRIGHT PROTECTION

No copyrighted audio files are redistributed or shared. The dataset is used in compliance with copyright laws.



USER PRIVACY

No personal or sensitive user data was collected. Complete privacy and data protection were ensured.

FEATURE PREPROCESSING



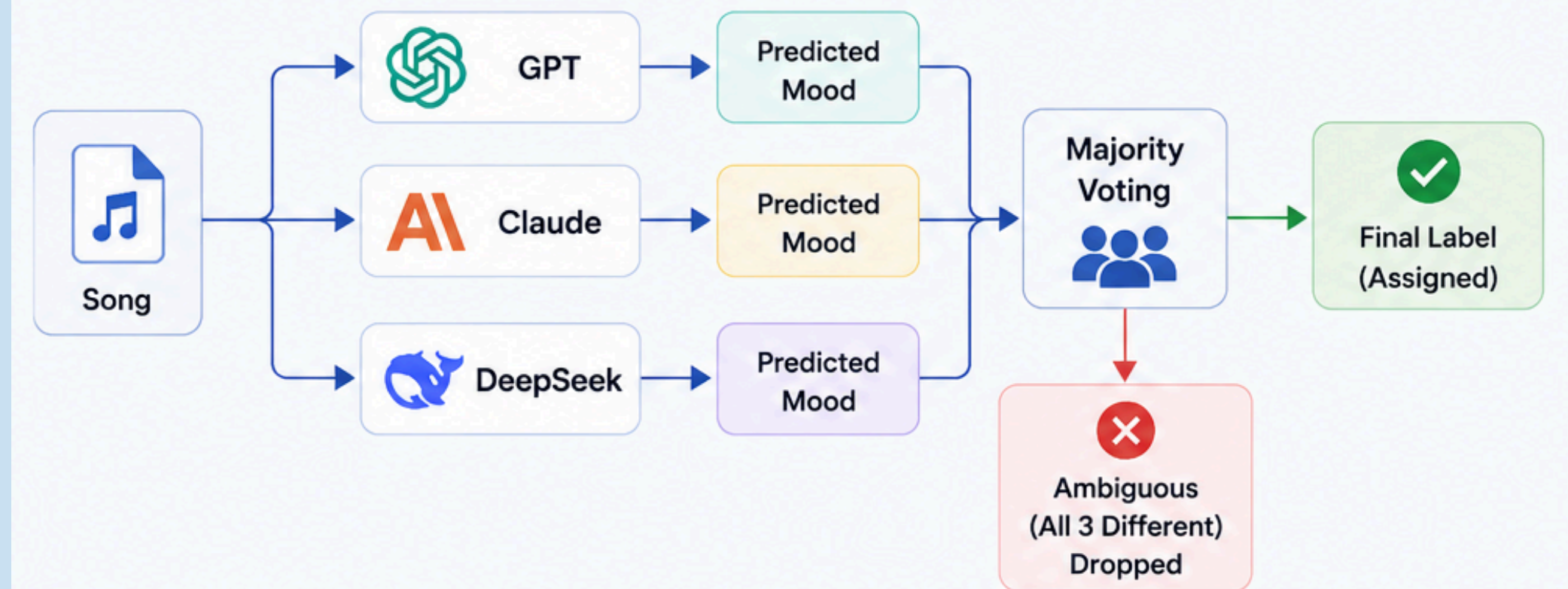
STEP 1- DATASET CREATION AND MOOD LABELS

Songs were collected across multiple mood categories:



GROUND TRUTH LABELING STRATEGY

Ground-truth mood labels were generated using a consensus-based approach with three LLMs.

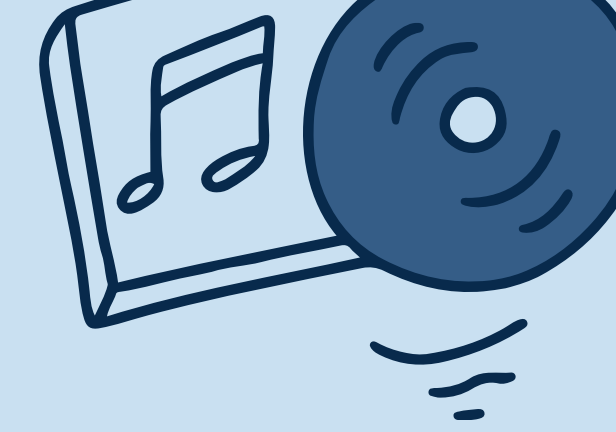


THIS HELPED:

Reduce labeling noise

Improve annotation consistency

Create cleaner decision boundaries for classification



FEATURE PREPROCESSING

STEP 2- FEATURE EXTRACTION AND MISSING VALUE HANDLING

Librosa Features

- MFCC coefficients
- Chroma features
- Spectral centroid
- Spectral contrast
- Zero-crossing rate
- Tempo and beat strength
- RMS energy and dynamic range

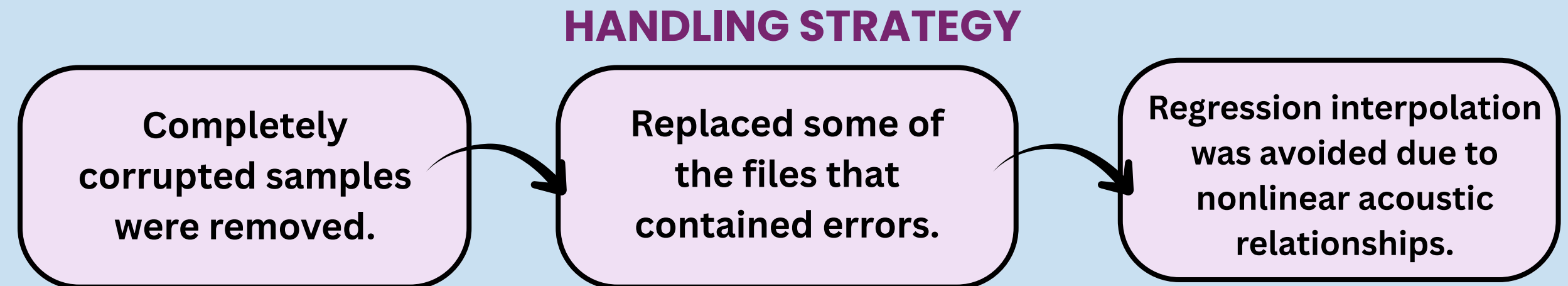
Essentia High-level Descriptors

- Valence
- Arousal

MISSING FEATURE HANDLING

Some songs produced incomplete feature vectors due to:

- Decoding failures
- Corrupted MP3 files
- Extraction-library errors

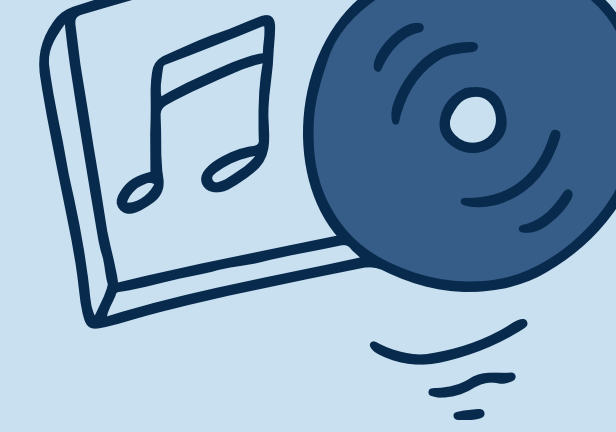


Key Observation:
MFCC analysis showed distinct mood-specific acoustic patterns-

➤ Romantic songs showed stronger harmonic richness.

➤ Party songs showed stronger rhythmic fluctuations.

➤ Excited songs showed smoother spectral behavior.



FEATURE PREPROCESSING

STEP 3: FEATURE SCALING AND DIMENSIONALITY REDUCTION

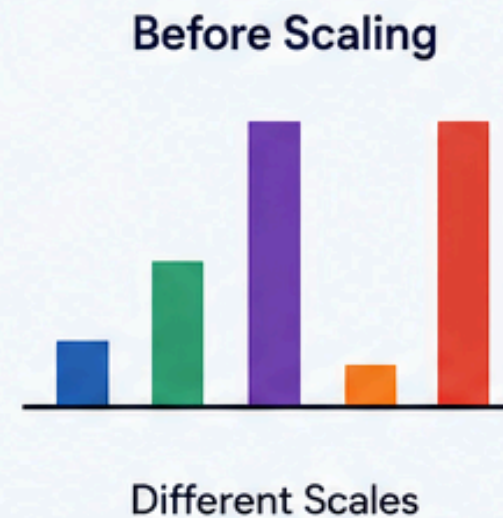


FEATURE SCALING

Extracted features existed on very different numerical scales.

Therefore:

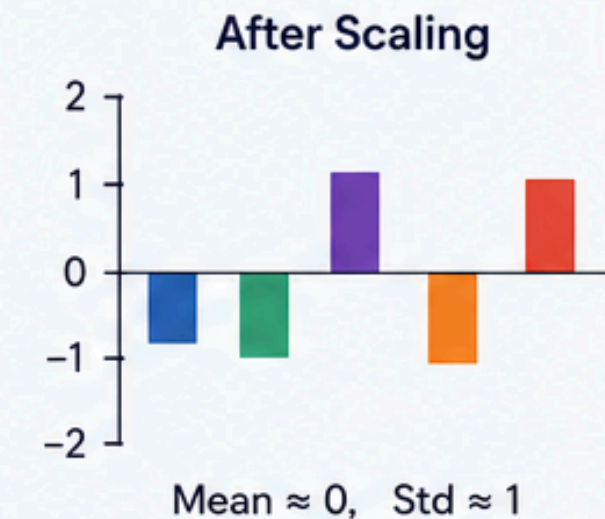
- StandardScaler was applied,
- features were normalized to mean 0 and standard deviation 1.



StandardScaler

$$Z = \frac{x - \mu}{\sigma}$$

$\mu = \text{mean}, \quad \sigma = \text{std}$



DIMENSIONALITY REDUCTION

Highly correlated and low-variance features were removed because:



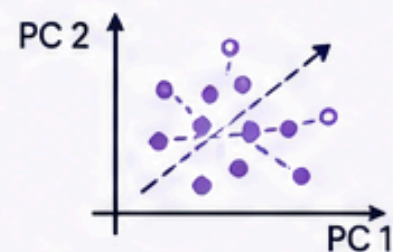
they contributed limited discriminative information



increased redundancy



could negatively affect model generalization



PCA was mainly used for exploratory dimensionality analysis and feature reduction support rather than fully replacing interpretable acoustic descriptors.

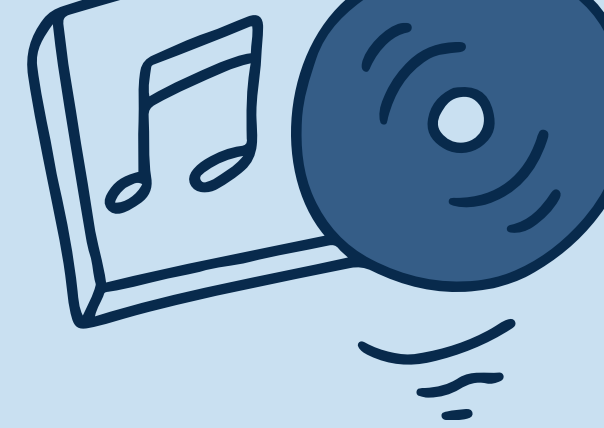
PURPOSE:

1 To reduce redundancy

2 To improve feature quality

3 To stabilize model learning

FEATURE PREPROCESSING

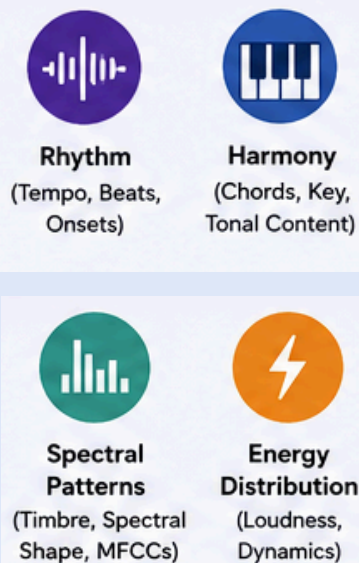


STEP 4: ESSENTIA FEATURE ANALYSIS & EMOTION MAPPING



HOW ESSENTIA WORKS

Essentia is an open-source audio analysis library that extracts high-level semantic and emotional descriptors from audio signals.



EMOTIONAL FEATURES EXTRACTED



Valence → Positivity or negativity of emotion
Range: -1 (negative) to +1 (positive)



Arousal → Energy or intensity level of emotion
Range: 0 (calm) to 1 (high energy)

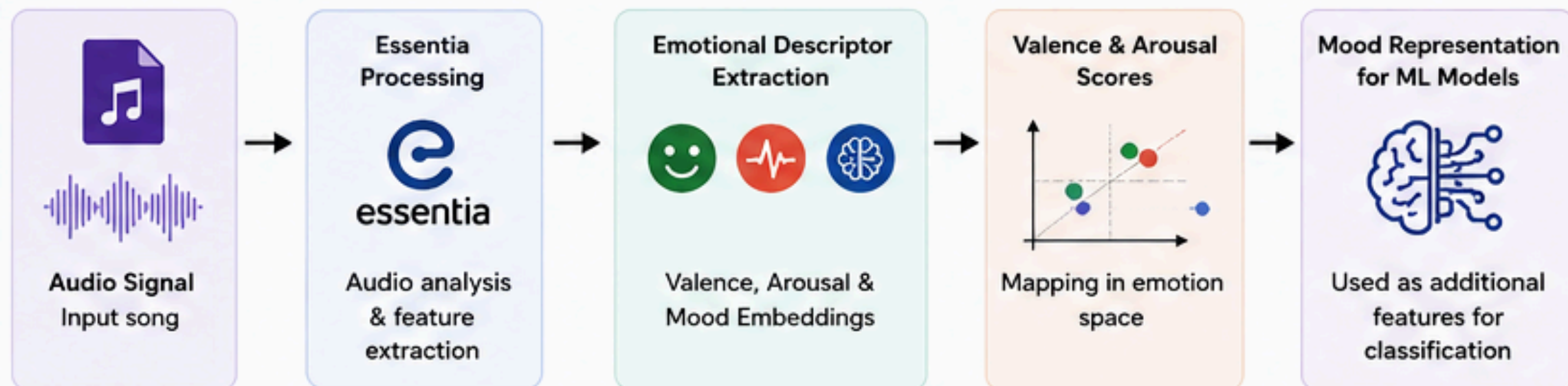
WHY ESSENTIA WAS USED

- Captures semantic emotional information beyond low-level acoustic features.
- Models psychological emotion representation in music.

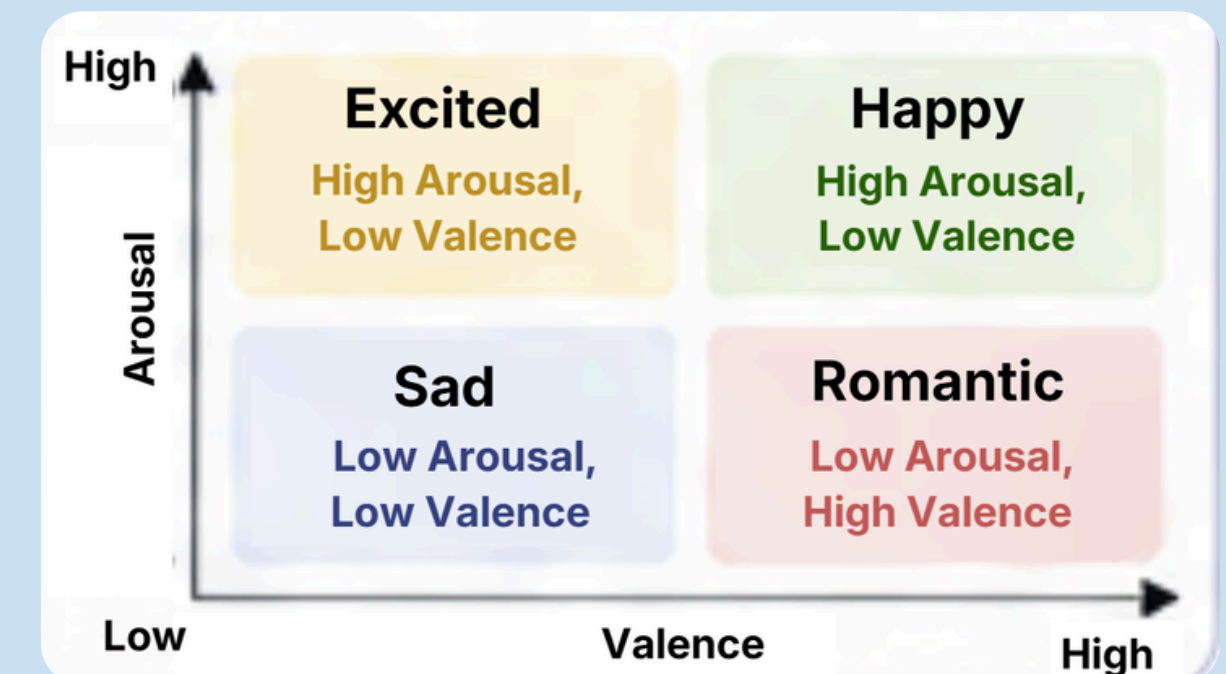


100+ HOURS OF COMPUTE TIME

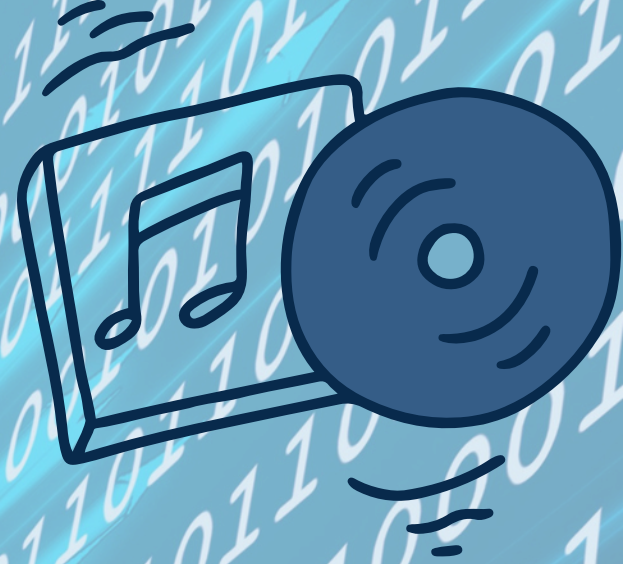
FEATURE INTEGRATION PIPELINE



VALENCE-AROUSAL EMOTION SPACE



ML METHDOLGY



1

Initial Baseline Approach: LOGISTIC REGRESSION

Why we started here

- Simple and interpretable multiclass baseline
- Standard benchmark for supervised classification
- Helped evaluate feature quality before complex models

Initial Results

- Accuracy plateaued around 43–44%
- Strong class confusion observed

Key Observations

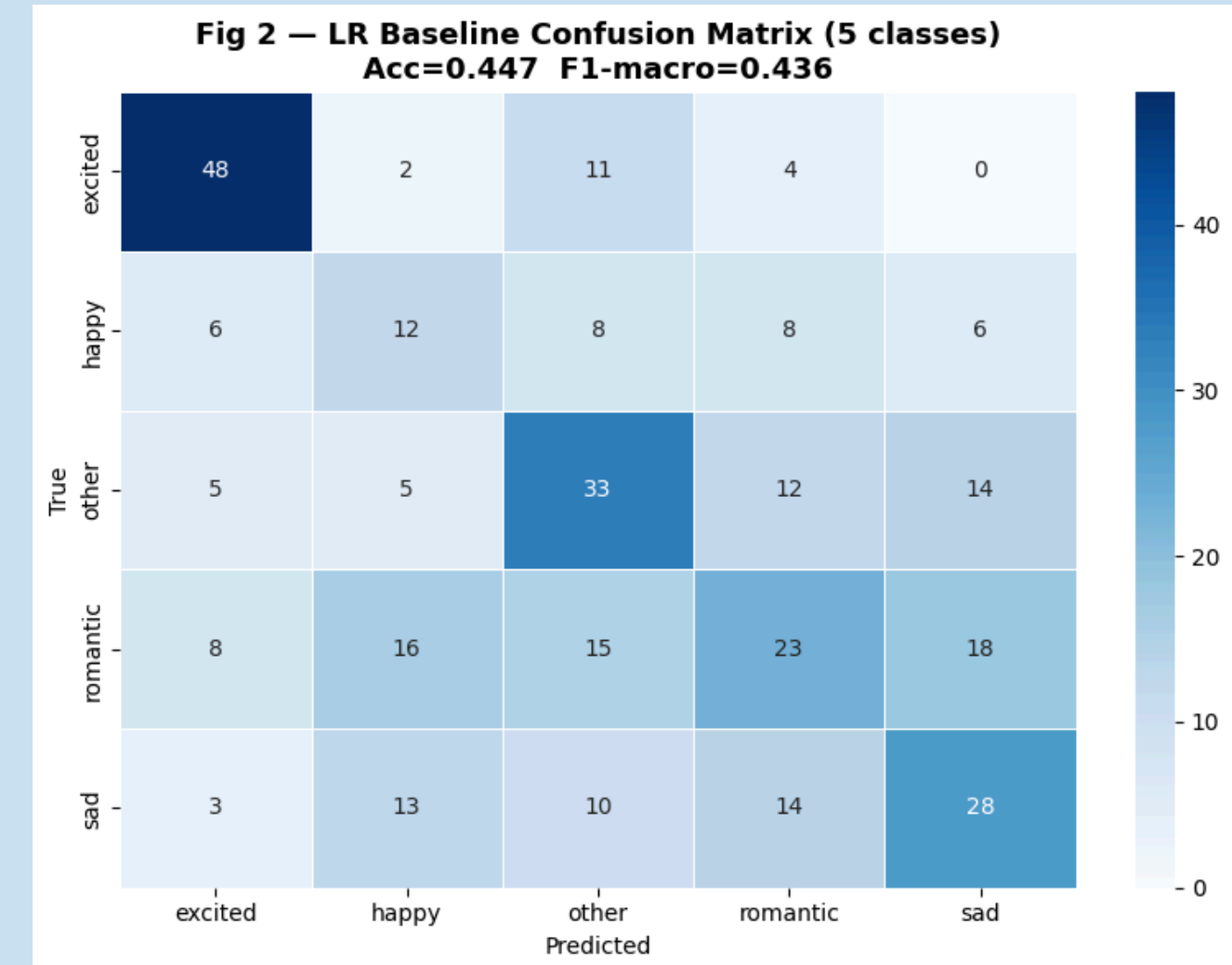
- “Happy” class poorly learned
- “other” class created noisy decision boundaries

```
=====
PHASE 1 – Logistic Regression Baseline (5 classes)
=====
Accuracy : 0.4472
F1-macro : 0.4360

      precision  recall  f1-score  support

excited    0.69    0.74    0.71     65
happy      0.25    0.30    0.27     40
other      0.43    0.48    0.45     69
romantic   0.38    0.29    0.33     80
sad        0.42    0.41    0.42     68

accuracy          0.45     322
macro avg         0.43     322
weighted avg      0.44     322
```



2

Dataset Refinement & Rebalancing

Step 1: Removed the "other" class

Reason:

- No clear acoustic identity
- Increased inter-class confusion

Result:

Accuracy improved significantly (~58%)

Step 2: Applied SMOTE/undersampling

Reason:

- "happy" had the lowest representation

Result:

- Only marginal improvement
- Suggested feature limitations rather than data imbalance

```

=====
PHASE 3 – LR without 'other' (4 classes)
=====
Accuracy : 0.5613 (was 0.4472, delta = +0.1141)
F1-macro : 0.5359 (was 0.4360, delta = +0.0999)

      precision    recall  f1-score   support

 excited      0.74      0.80      0.77      65
   happy      0.24      0.30      0.27      40
  romantic      0.58      0.41      0.48      80
     sad      0.59      0.66      0.62      68

 accuracy          0.56      253
 macro avg          0.54      253
 weighted avg          0.57      253
  
```

Phase 3: Impact of Removing 'other' Class

Fig 3a – Baseline (4 shared classes shown)
Acc=0.447 F1=0.436

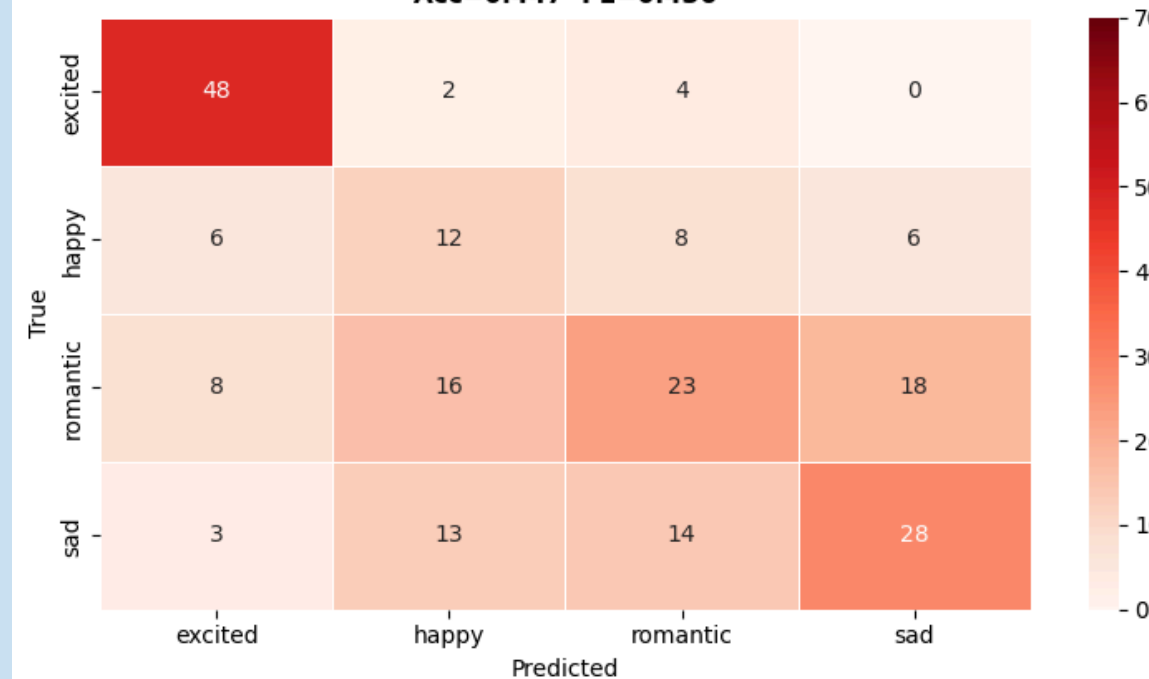
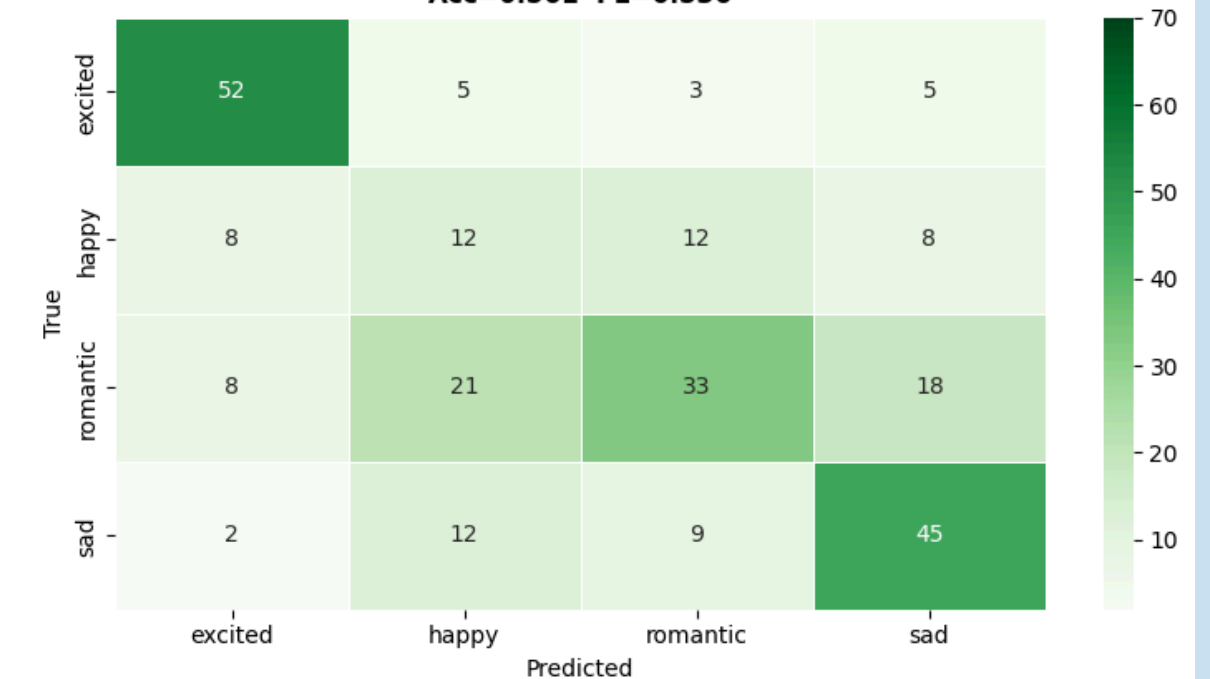


Fig 3b – Remove 'other' (4 classes)
Acc=0.561 F1=0.536



Experimenting with Emotional Features

Hypothesis

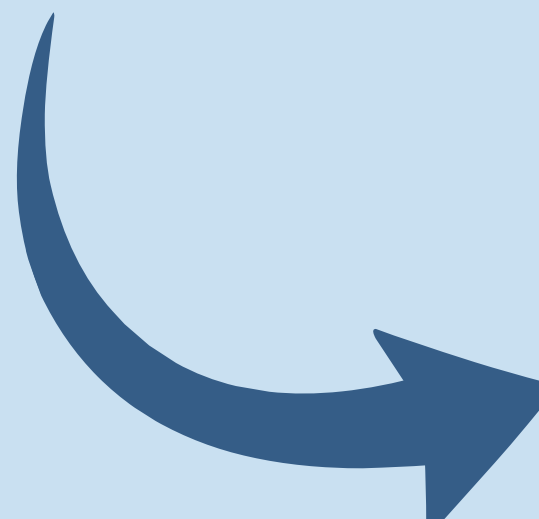
- Adding Essentia valence/arousal features would improve emotional understanding and mood classification.

Observation

- Model accuracy decreased after integrating these features.
- High-level emotional descriptors introduced additional noise.

Interpretation

- Essentia models are primarily pretrained on Western music datasets.
- These features showed poor generalization for Hindi/Hinglish songs due to cultural and acoustic differences.



4

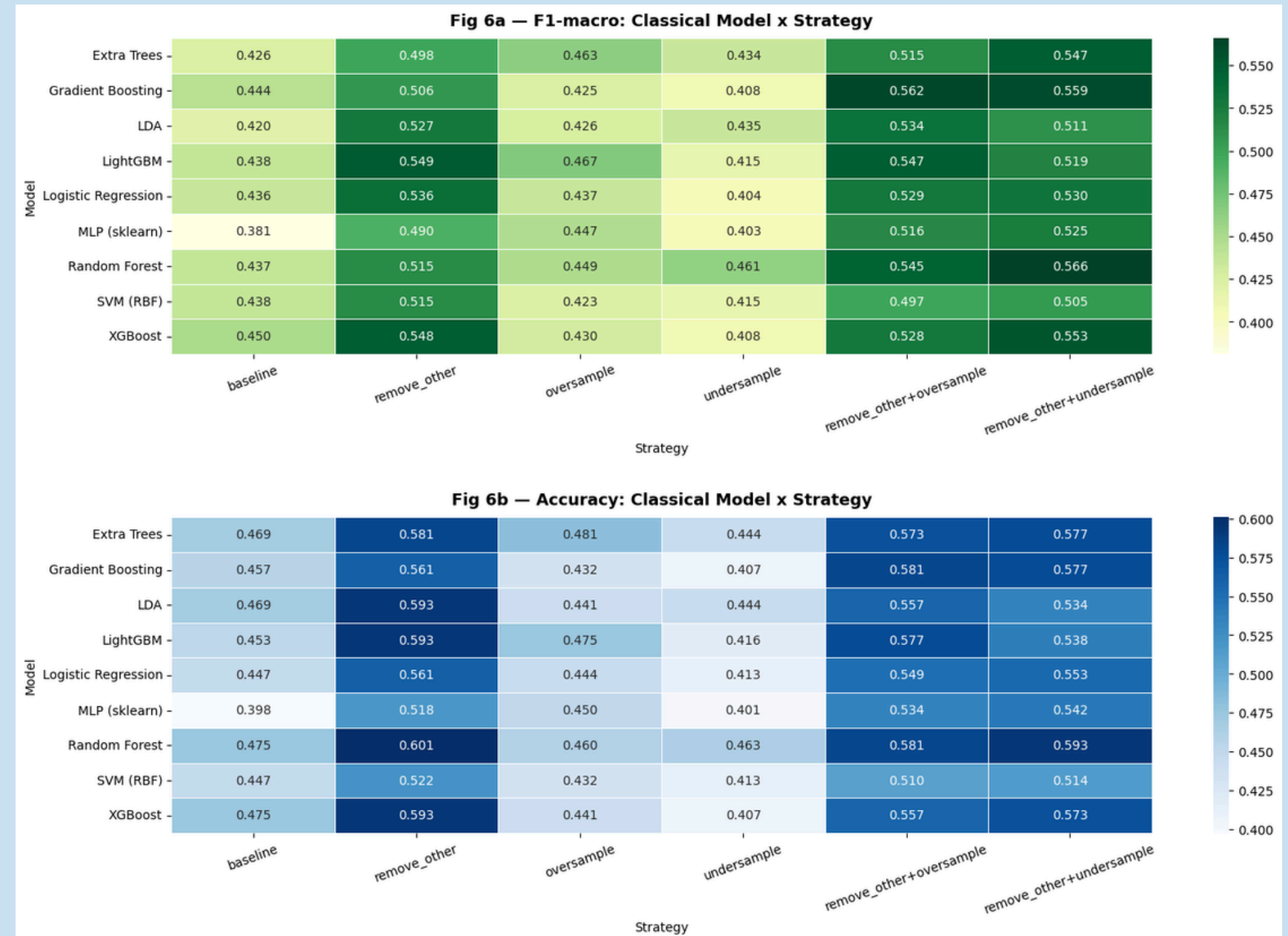
Exploring Non-Linear Models

MODELS TESTED

- 1 Random Forest
- 2 XGBoost
- 3 SVM
- 4 Gradient Boosting
- 5 LDA

Observations

- Removing the “other” class gave the largest performance improvement across all models.
- Random Forest achieved the best classical ML accuracy (~60.1%).
- Sampling techniques alone showed limited improvement, suggesting feature complexity was the key limitation.



5

Experimentation with DL Model

Why Deep Learning?

- Classical ML models plateaued around 58–61% accuracy.
- Audio features showed complex non-linear relationships.
- Stronger feature representation learning was needed.

Why TabTransformer?

- Uses self-attention to learn feature interactions.
- Better suited for structured tabular audio features.

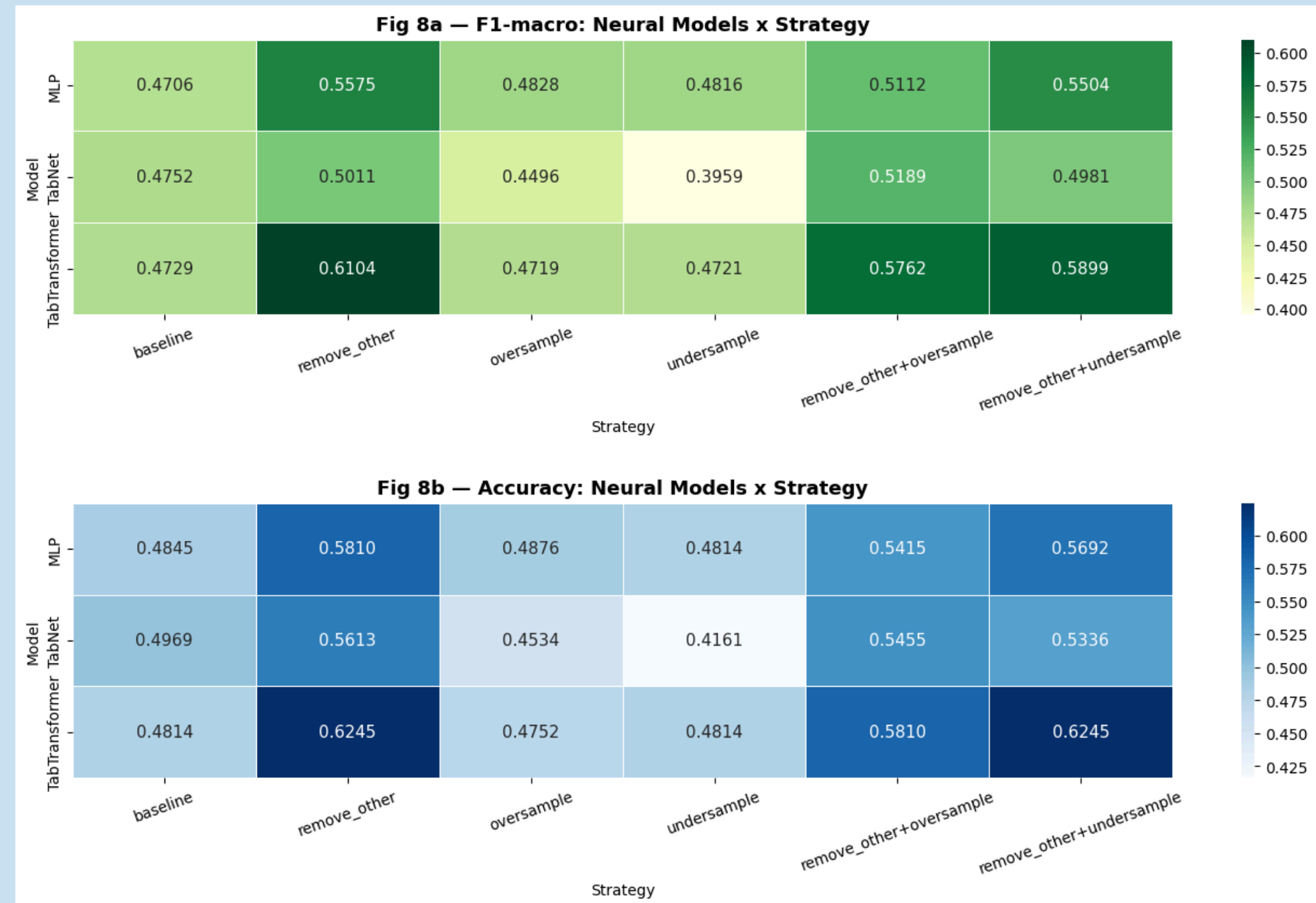
Advantages

- Learned complex MFCC/chroma relations
- Better feature representation
- Improved generalization

Final Performance

- Accuracy: 0.6245
- F1-score: 0.6104

Final Architecture: TabTransformer



KEY CHALLENGES

No suitable Hindi/Hinglish dataset available

Required manual collection and labeling of songs.

Spotify & Genius API limitations

Lyrics and audio metadata could not be extracted directly due to deprecated/restricted APIs.

Difficulty in Hindi lyric extraction

Speech-to-text models performed poorly on Hindi songs with music and mixed vocals.

Feature overlap between moods

Romantic and sad songs shared similar acoustic patterns, causing confusion.

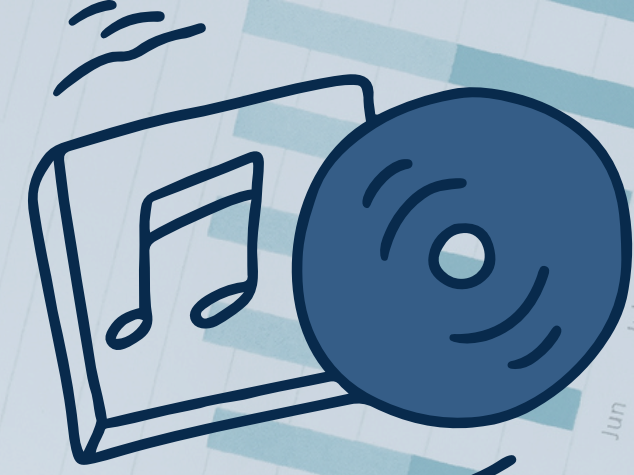
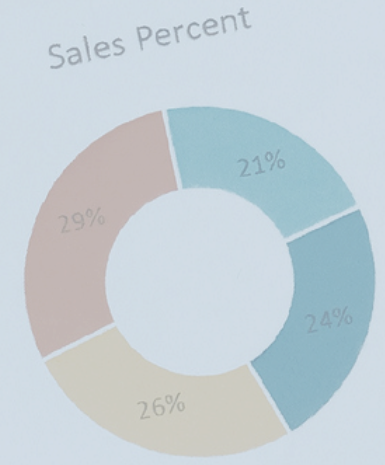
Feature engineering challenges

Initial audio features were not sufficiently discriminative for mood prediction.



PERFORMANCE METRICS

Month	Sales 1	Sales 2	Sales 3	Sales 4
Jan	32453	36754	50565	45674
Feb	34677	43456	53211	56333
Mar	41345	56755	53422	58423
Apr	45242	56978	54568	68211
May	43546	57785	56437	69424
Jun	53654	56353	57675	75234
Jul	56743	58976	59646	76111
Aug	56783	58976	67454	76533
Sep	59564	58976	67454	79535
Oct	67432	60876	78454	79535
Nov	68422	64532	79535	80353
Dec	70478	68953	80535	81556
Total	559867	620394	691502	767387
Percent	21.3%	23.3%	26.38%	28.89%



PERFORMANCE METRICS

Metric / Observation	Why It Validates Our System
Random Forest Accuracy	Achieved approximately ~61% accuracy, showing significant improvement over baseline Logistic Regression (~43–44%) and better overall mood prediction capability.
F1-Score	Improved F1-score indicated better handling of overlapping emotional classes such as sad and romantic, making predictions more balanced and reliable.
Precision	Higher precision reduced incorrect mood assignments, improving the quality of mood-based recommendations and playlist generation.
Recall	Improved recall helped the model identify a larger proportion of actual mood classes, especially underrepresented categories like happy.
Confusion Matrix Analysis	Showed reduced confusion between emotional categories after dataset refinement and advanced ML modelling, validating improved emotional understanding.



BENCHMARK COMPARISON

Research Paper	Metrics Achieved in Research Paper	Metrics Achieved by Our Model
Lyrics-Based Music Genre Classification Using HAN	Accuracy: 46.42%	Accuracy: ~61%
Enhancing Accuracy and Performance in Music Mood Classification	Accuracy: 56.8% F1-Score: 0.541 Precision: 0.548 Recall: 0.568	Accuracy: ~61% F1-Score: ~0.60 Precision: Improved over baseline LR Recall: Better balanced across moods
Joint Sentiment Analysis of Lyrics & Audio	Accuracy: ~60% F1-Score: Improved weighted F-score Recall: Better valence/arousal detection	Accuracy: ~61% F1-Score: ~0.60 Recall: Balanced mood-wise performance

DEPLOYMENT AT PLAKSHA



Student Wellness and Mental Health Support

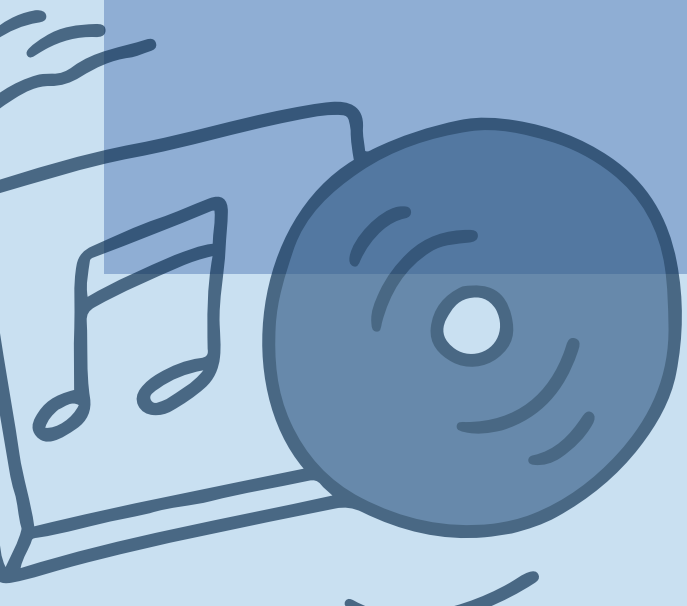
- Mood-aware music recommendation system
- Stress-relief and relaxation playlists
- Personalized music suggestions based on emotional state

Smart Campus Entertainment and Events

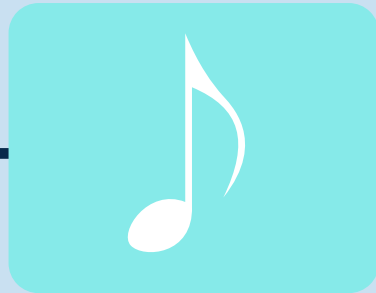
- Automatic playlist generation for campus events
- Mood-based background music for cafes or common areas
- Personalized music experience during student activities

AI-Based Research and Personalization Platform

- Real-time Hindi music mood classification
- Integration with student apps or wellness platforms
- Future expansion to multilingual and genre-adaptive recommendation systems



DEPLOYMENT CHALLENGES AT SCALE



HANDLING DIVERSE MUSIC PREFERENCES

Mood perception varies across users, languages, and cultural backgrounds, making large-scale personalization difficult.



DATASET EXPANSION AND MAINTENANCE

Continuously updating the system with new Hindi/Hinglish songs and accurate mood labels would require ongoing data collection.



FEATURE INCONSISTENCY ACROSS MUSIC STYLES

Acoustic patterns may vary significantly across genres, remixes, and regional music formats, affecting prediction consistency.



MODEL GENERALIZATION CHALLENGES

Models trained on limited datasets may struggle when exposed to unseen artists, languages, or evolving music trends.

Thank You

