International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Matrix Factorization Using LightFM for a Music Recommendation System Based on Emotional and Listening Behavior Awareness

Anindita Putri Dayati[1], Indriani[2]

Abstract

This paper summarizes an emotion-aware hybrid music recommendation approach that combines users' emotional listening characteristics and implicit behavioral signals to enhance personalization under sparse feedback conditions. We construct enriched user profiles by aggregating audio-derived emotional features such as valence and danceability, modeling genre preferences through one-hot encoding, and capturing engagement behavior via skip-rate statistics, followed by systematic preprocessing including outlier removal and normalization. Using standardized emotional features, we apply K-means clustering to assign interpretable mood contexts (e.g., happy, energetic, calm, and sad), which are then incorporated as user-aware signals in a hybrid LightFM matrix factorization model optimized with the WARP-kos loss for ranking-based recommendation. Experimental evaluation demonstrates that the proposed model achieves a Precision@10 of 0.6209, indicating that more than six out of ten recommended tracks are relevant, and a Recall@10 of 0.4663, meaning that approximately 47% of all relevant items are successfully retrieved within the top-10 recommendations. These results highlight the model's ability to balance accuracy and coverage while outperforming traditional collaborative and content-based baselines, thereby confirming that integrating emotional context with behavioral data significantly improves the effectiveness of personalized music recommendation systems.

Keywords:
Music Recommendation, LightFM, Hybrid Matrix Factorization, Emotion State, Listening Behaviour, K-Means

1. Introduction

Music surpasses entertainment. Deeply intertwined with human emotions, it serves as an impressive tool for achieving peace of mind, relieving stress, and regulating mood [1], fostering creativity [2], and evoking emotions that encourage kindness and cooperation [3]. In society, shared enthusiasm for a genre or an artist often brings people together in fan communities, forging strong bonds and providing mutual support for both the music and its members. These effects highlight music's profound impact on individual psychological well-being and social connections.

The digital era has fundamentally transformed music consumption. Physical media, such as CDs and DVDs, have largely been supplanted by music streaming platforms, whose convenience, accessibility, personalized discovery, seamless device integration, and other benefits have driven their explosive global growth and established them as the dominant distribution channel. However, these abundances of content create challenges, particularly in discovering music that matches a user's current emotional state. This underscores the need for effective assistance of a music recommendation system (MRS).

Corresponding Author: Indriani (Indriani.st@widyatama.ac.id)
1 Anindita Putri Dayati, Widyatama University, anindita.dayanti@widyatama.ac.id
2 Indriani, Widyatama University, indriani.st@widyatama.ac.id

Traditional music recommendation systems (MRS) rely primarily on collaborative filtering (CF) or content-based filtering (CBF), matching users based on historical reference or song metadata. These approaches often neglect critical psychological and contextual influences on musical preferences, such as emotional states and listening behaviors [4], [5]. In response, modern music recommendation systems have evolved toward context-aware and emotion-aware approaches. For instance, the emotion state transition model (ESTM) links emotion, context, and music features to tailor recommendations to the user's emotional state [6]. Wang & Chou [7] proposed incorporating both emotional and contextual factors (time, location, or activity), improving relevance and user satisfaction. Rozhevskii [8] combined psychological traits with transient emotions for greater precision. Further, advancements exploit effective signals such as facial expressions, voice, and text to infer valence and arousal, then match songs using deep learning models such as DRViT and InvNet50, enabling more nuanced recommendations [9]. Recent studies have continued to push boundaries, introducing heterogeneity-aware deep Bayesian Networks for emotion modeling [10], real-time emotion context integration for enhanced user experience [11], cross-cultural emotion benchmarks to improve generalization across diverse listeners [12], and an emotion-adaptive framework using natural language inputs for dynamic mood alignment [13].

In music recommendation systems, a two-stage pipeline is followed: explicit emotion detection, followed by recommending music [9], [14]. This separation can propagate errors across stages. Recent research addresses this by bypassing explicit emotion labeling, instead learning directly from behavioral interaction data–such as keystroke history, clicks, and navigation patterns–which can serve as implicit emotional indicators [15], [16]. These methods enhance the level of emotional personalization while remaining practical and unobtrusive, with newer work leveraging multimodal signals and adaptive architecture to capture dynamic mood shifts from listening patterns [13].

Sequential listening behaviors, such as skipping songs, replaying them, or listening durations, offer rich, dynamic signals of evolving preferences [17], [18]. Unlike static metrics such as ratings and listener counts, these implicit cues capture real-time mood shifts and contextual habits in an environment where preferences fluctuate, and songs are often revisited. Including emotional awareness and listening behavior awareness in MRS yields significant benefits, surely elevating performance and user outcomes. The role of an emotion-aware system incorporated with the user's transient mood would lead to higher user satisfaction, improved engagement, enhanced retention, and enriched emotions, such as mood regulation, relaxation, and therapeutic effect [19]. Meanwhile, listening behavior awareness, implicit feedback, enabling reliable personalized, instantaneous behavioral tracking, disregarding explicit input, mitigating cold-start issues, and capturing dynamic signals that static methods overlook[20].

Hybrid matrix factorization has emerged as a powerful method for building a robust recommendation engine that integrates both emotional and behavioral data on a large scale, in a sparse environment typical of music streaming. Among available models, LightFM is noticeable for several compelling reasons. To begin with, it is explicitly designed to handle implicit feedback datasets, prevalent in modern platforms where explicit ratings are rare, but behavioral signals using the Weighted Approximate Rank Pairwise loss function to optimize precision ranking directly relevant to user satisfaction [21], [22],[23]. In addition, LightFM incorporates high-dimensional auxiliary feature side information (e.g., emotional clusters, genre embedding, audio features like valence and danceability, and behavioral metrics) into latent representation, effectively bridging collaborative, content-based, and contextual modelling while mitigating cold-start problems for new users or tracks. Furthermore, its complexity and efficiency training on sparse matrices enable scalability to millions of users and items, making it practical for real-world deployment. Lastly, across the recommendation music task through benchmarks, LightFM demonstrates performance over CF or CBF in handling heterogeneous data sources, which leads to more accurate and diverse recommendations. Conclusion: these senses make LightFM well-suited for emotion-awareness that is accompanied by rich multimodal features.

This study proposed an emotion-aware music recommendation system built on the LightFM hybrid matrix factorization framework. The model enriches the latent representation space by integrating user-specific mood clusters, genre preferences, and listening patterns. Unlike methods reliant on explicit emotion detection, our approach uses implicit affective signals from behavior (e.g., skip frequency, session continuity, and play duration) alongside audio-based features (e.g., valences and danceability). By jointly modeling emotional and behavioral contexts, the system delivers recommendations that are personalized, contextually relevant, and emotionally aligned with the user's listening, fostering greater satisfaction and potential well-being benefits.

2. Related Works

A music recommendation system (MRS) leverages artificial intelligence to generate personalized music suggestions by analyzing a user's explicit and implicit preferences. Its system has become a critical component of leading streaming services, including YouTube, Spotify, Apple Music, and so on, by enabling users to efficiently discover music that suits their individual tastes or situational needs.

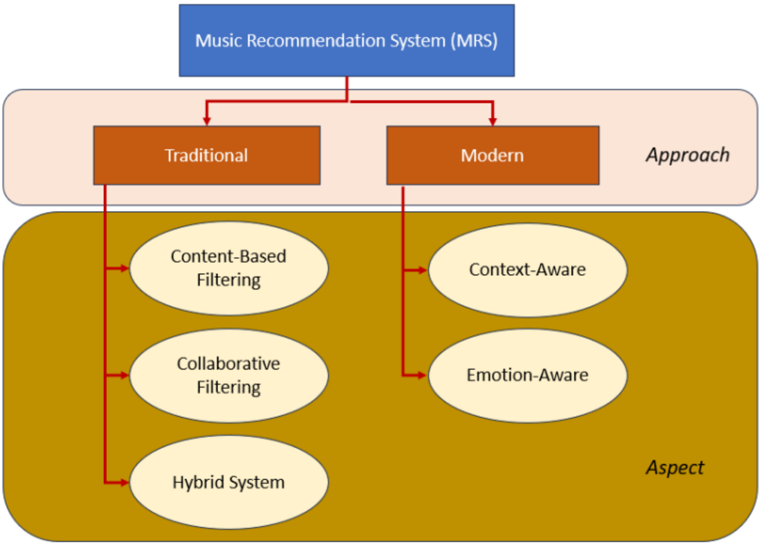


Fig. 1. MRS Categories

Approaches for music recommendation systems, as illustrated in Fig. 1, are categorized into two main groups: traditional and modern. Traditional approaches encompass Content-Based Filtering (CBF), Collaborative Filtering (CF), and a Hybrid System. In contrast, modern approaches include Context-Aware or Emotion-Aware recommendation systems. Content-Based Filtering provides a user's historical preferences from the user's personalized experience based on the similarity of song features, such as genre [24], tempo [18], [24], [25], [26], lyrics [27], [28], or instrumentation [29], [30] to generate music recommendations. However, it has adverse effects, such as promoting overspecialization, creating a filter bubble that hinders the discovery of serendipitous and novel content.

Unlike Content-based methods, Collaborative Filtering relies on behavior patterns such as play count, likes, or skips without analyzing the content itself as user references. It predicts a user's interest by identifying other users with similar taste patterns [31]. Methodologically, the CF is realized through two approaches: first, the memory-based (neighborhood) method, which calculates the entire user-item interaction matrix, including User-User Collaborative Filtering [32], and Item-Item Collaborative Filtering [33]. The last one, the model-based method, which learn a predictive model from interaction data, is a method well known as Matrix Factorization [34] and Advanced variants like SVD++ [35]. Collaborative Filtering has to deal with the cold-start problem, where it cannot reliably recommend new music or new users due to a lack of interaction history [36]. Furthermore, data sparsity in large music catalogs can hinder accurate similarity calculation [4]. To mitigate these limitations, the music recommendation system has evolved toward a hybrid model that combines CBF and CF to leverage their respective strength [37], and deep learning techniques have been applied, such as Neural Collaborative Filtering [38] and Graph Neural Networks [39].

In a modern music recommendation system, to provide highly personalized recommendations by adapting not only the user's long-term preferences, but also their immediate context (time, location, activity), and emotional state or behavior requires integrating multiple modalities of information to construct a richer understanding of both music and users. These systems are suggesting individual tracks but also capable of recommending playlists, i.e., coherent sequences of songs that align with the user's evolving mood and situational needs [40]. To enhance user satisfaction, modern MRS further strives to balance accuracy with diversity, thereby reducing the risk of filter bubbles while still ensuring relevance [41], and is competent in the emphasis of scalability and real-time adaptation to handle user interactions at a large scale. Collectively, these aspects support modern music recommendation systems to deliver more engaging, context-aware, and responsive listening experiences.

3. Proposed Method

This study applies a hybrid recommendation approach that combines collaborative filtering and content-based filtering within the LightFM framework to address the limitations of single-method recommender systems. Collaborative filtering leverages implicit user–item interaction data to capture collective listening preferences, while content-based filtering integrates item-level audio features to model intrinsic characteristics of music tracks. In this work, we enrich the item representations by incorporating audio-derived emotional features, such as valence, arousal, and mood-related descriptors extracted from signal processing or pretrained audio analysis models. At the same time, we model user listening behavior patterns, including skip rate, session duration, repeat frequency, and genre affinity. It is to capture short-term engagement and long-term preference tendencies. By embedding both emotional cues and behavioral signals into the latent factor space, the proposed method enables LightFM to learn more expressive user and item embeddings that reflect not only co-listening patterns but also affective alignment and contextual relevance.

In this study, this hybrid design allows the recommendation system to adapt to dynamic user states and mitigate common challenges such as cold-start and preference drift. Emotional features provide semantic grounding for new or sparsely interacted tracks, while behavioral indicators help infer user intent and satisfaction beyond simple play counts. LightFM's factorization model naturally fuses these heterogeneous signals by learning shared latent dimensions that jointly explain interaction data and content features, resulting in recommendations that are more personalized and emotionally coherent. As a result, the system can recommend music that better matches users' current moods and listening contexts, rather than relying solely on historical popularity or similarity. This emotionally and behavior-aware representation enhances recommendation accuracy, improves user engagement, and supports more human-centered music discovery experiences.

In this paper, we adopted LightFM as a hybrid recommendations model that applies matrix factorization by incorporating user and item side information [23]. The predicted interaction score $\hat{r}_{ui}$ between user u and item i is the dot product of the latent representation:

$$\hat{r}_{ui} = f\,(q_u \cdot p_i + b_u + b_i) \qquad (1)$$

where q_u is the latent vector for the user u,, p_i is the latent vector for the item i, meanwhile b_u and b_i refers to user bias and item bias. So, the latent representations are linear combinations of feature embedding:

$$q_u = \sum_{f \in F_u} e_f^{(u)},\ p_i = \sum_{g \in F_i} e_g^{(i)} \qquad (2)$$

where F_u refers to a set of features for the user u, $e_f^{(u)} \in \mathbb{R}^d$ for latent embedding for user features f, while $e_f^{(u)} \in \mathbb{R}^d$ for the similarity of item features g.

4. Experimental Setup

In this section, explain how the experiment was conducted and summarize the data taken during the experiment with a series of processes of an emotion-aware and listening behaviour for a recommendation system using the LightFM hybrid matrix factorization framework. The overall research workflow, as illustrated in Fig. 2, employs an experimental methodology comprising a structured sequence of stages: data collection, preprocessing, emotional clustering, user-item interaction modeling, determining item features, and training and evaluation using the LightFM hybrid matrix factorization framework.

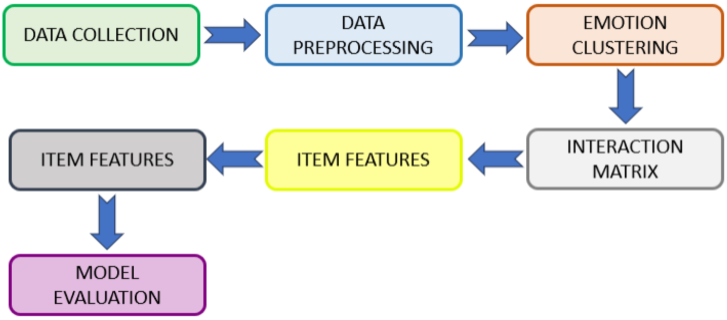


Fig 2. Research Flowchart

A. Data Collection

In this study, we utilize multiple complementary datasets to support the development of a hybrid emotion-aware music recommendation system that integrates behavioral, contextual, and affective information. Specifically, we employ three publicly available datasets sourced from Kaggle to capture different aspects of user–music interactions. First, we use the Spotify Streaming History Dataset, which contains approximately 149,860 records of user listening activity, including track titles, artist names, timestamps, and contextual metadata. We apply this dataset to construct the user–item interaction matrix using implicit feedback signals such as listening duration, play frequency, and skip behavior, which reflect user engagement and preference strength more accurately than explicit ratings. These interaction patterns form the foundation for collaborative filtering within the recommendation model.

In addition, this paper utilizes the Spotify Playlists Dataset and the Spotify Tracks Dataset to enrich the recommendation process with semantic and emotional content features. The Spotify Playlists Dataset comprises over 12 million playlist entries, representing user-curated song groupings that implicitly encode thematic, genre-based, and mood-driven listening preferences. We adopt this dataset to model higher-level listening patterns and contextual associations between tracks. Furthermore, we use the Spotify Tracks Dataset, which provides audio and metadata features for more than 114,000 tracks, including valence, energy, danceability, tempo, and genre information. These attributes enable content-based filtering and emotional feature extraction, allowing the system to align recommendations with users' affective states. All datasets are provided in CSV format and processed using Python libraries such as pandas for data loading, cleaning, merging, and feature engineering. By integrating these datasets, we apply a unified framework that captures behavioral dynamics, contextual cues, and emotional characteristics to support more personalized and emotionally aware music recommendations.

B. Data Preprocessing

In this study, we apply a comprehensive data preprocessing pipeline to enrich and standardize raw data obtained from the Spotify Streaming History, Playlists, and Tracks datasets, ensuring consistency, quality, and suitability for recommendation modeling. We first clean and integrate the datasets by aligning track identifiers and removing incomplete or inconsistent records. To address noise and extreme values that could bias the learning process, we detect and remove outliers using the Interquartile Range (IQR) method across key numerical attributes, including valence, energy, tempo, and total play duration. We then normalize all numerical features using Min–Max scaling, which maps values into a common range and prevents features with larger magnitudes from dominating the latent factor learning and similarity computations. This normalization step is particularly important for hybrid models such as LightFM, where heterogeneous features are jointly embedded into a shared latent space.

Furthermore, this paper constructs user-level feature profiles by aggregating behavioral and audio characteristics to capture stable preference patterns. We compute the mean of audio features per user to represent dominant emotional tendencies, identify the most frequent genre using the statistical mode to reflect genre affinity, and calculate the average skip rate as an indicator of listening satisfaction and engagement. To enable seamless integration with the recommendation model, we transform categorical genre attributes into numerical representations using one-hot encoding. This preprocessing stage establishes the core foundation for building the user–item interaction matrix and feature representations by effectively combining user behavior, contextual listening patterns, and emotional audio descriptors. As a result, the processed data supports more robust latent factor learning and improves the model's ability to generate emotionally and behaviorally aligned music recommendations.

C. Recommender System

In this study, we apply user profile clustering to explicitly capture the emotional context underlying individual listening behavior. We utilize key audio attributes as primary indicators of emotional perception, as these features strongly correlate with mood and affective response to music. We combine these emotional attributes with genre indicators to form a comprehensive clustering dataset that reflects both affective and musical preferences. To ensure that all features contribute equally to the clustering process, we standardize the dataset using z-score normalization. We then adopt the K-Means algorithm to group users based on similarity in emotional and musical profiles. To determine the optimal number of clusters, we apply the elbow method in conjunction with the silhouette score, enabling a balance between cluster compactness and separation. Each resulting cluster is qualitatively labeled with mood categories such as *happy*, *sad*, *calm*, and *energetic*, based on average valence and danceability scores, and these mood labels are incorporated as high-level user features in the recommendation model.

To model user preferences effectively, this paper constructs an interaction matrix using implicit feedback derived from listening behavior. We aggregate user–track interactions by summing total play duration for each track per user, which reflects the strength of engagement beyond binary consumption. We then normalize these interaction values using Min–Max scaling to map them onto a 0–1 range, ensuring consistent weighting across users and tracks. User IDs and track IDs are mapped into a sparse matrix representation using the Coordinate (COO) format, which is computationally efficient and well-suited for the LightFM framework. In addition to the weighted interaction matrix, we also generate a binary interaction matrix, where a value of 1 indicates any non-zero interaction. This binary representation captures basic user interest and serves as the primary interaction signal for implicit feedback modeling, allowing the system to learn preference rankings rather than explicit ratings.

Furthermore, we enrich each music item with content-based features to support the hybrid recommendation strategy. We derive numerical audio descriptors such as valence, energy, danceability, tempo, and instrumentalness. It discretizes them into categorical bins (e.g., low, medium, high) to improve interpretability and reduce noise. Tempo is categorized separately into slow, medium, and fast classes to reflect rhythmic characteristics more clearly. We also one-hot encode track genres to provide contextual and stylistic information that complements audio-based features. All categorical features associated with each track are combined and formatted as feature tuples compatible with LightFM. This process produces an item feature matrix that links each track ID to its corresponding content attributes, enabling the model to integrate collaborative signals with rich item-level semantics during training.

We train the hybrid recommendation model using the LightFM framework with the WARP-kos loss function, which is specifically designed for ranking tasks in implicit feedback scenarios. We configure the model with a learning rate of 0.07, 160 latent components, and train it for 1,200 epochs using four parallel threads to ensure convergence and computational efficiency. Training inputs include both the interaction matrix and the item feature matrix, allowing LightFM to jointly learn collaborative and content-based representations. To evaluate model performance, we apply Precision and Recall metrics, which are appropriate for implicit recommendation settings and provide a balanced assessment of accuracy and coverage. We implement a time-based train–test split, where each user's most recent interaction is reserved for testing, to simulate realistic recommendation scenarios and assess the model's ability to predict future preferences rather than memorize past behavior.

5. Result and Analysis

The preprocessing stage produced a clean, integrated dataset of user interactions, track metadata, and audio features, ready for clustering and recommendation modeling. To ensure diversity and computational feasibility, 5,500 unique users were randomly sampled. This sample was merged with listening history and track metadata, resulting in a comprehensive dataset capturing user behavior and song characteristics.

To maintain data quality, key features that had missing values, such as valence, energy, danceability, tempo, and total playing duration, were removed. Then, to exclude extreme values that could distort the feature distribution from the dataset, the dataset was filtered using the Interquartile Range (IQR). Remaining numerical features were normalized via Min-Max Scaling to the [0,1] range, ensuring the consistency of the feature scale does not affect the stability of the model. The user profiles were constructed by aggregating the average value of emotional audio features for each user. In addition to audio-based, behavioral signals were also extracted, resulting in: (1) The most frequently listened genres calculated using mode aggregation per user, and (2) The average skip rate calculated based on user interactions, providing insights into the level of engagement. Categorical genre labels were one-hot encoded and then combined into the user profile dataset. The resulting hybrid representation is enriched with emotional tendencies (e.g., average valence, dancing ability) and behavioral context (e.g., genre preference, skipping behavior), serving as the foundation for personalized clustering and recommendations in the next stage.

Post-processing, 3,727 complete user profiles were clustered based on emotional listening patterns using the K-Means algorithm, applied to two standardized features: valence (emotional positivity) and danceability (rhythmic movement). The optimal number of clusters was determined using both the elbow method and silhouette score analysis. As shown in Fig. 3, the silhouette score reaches its peak at K = 15, with a value of 0.5325, indicating an excellent degree of cluster separation.

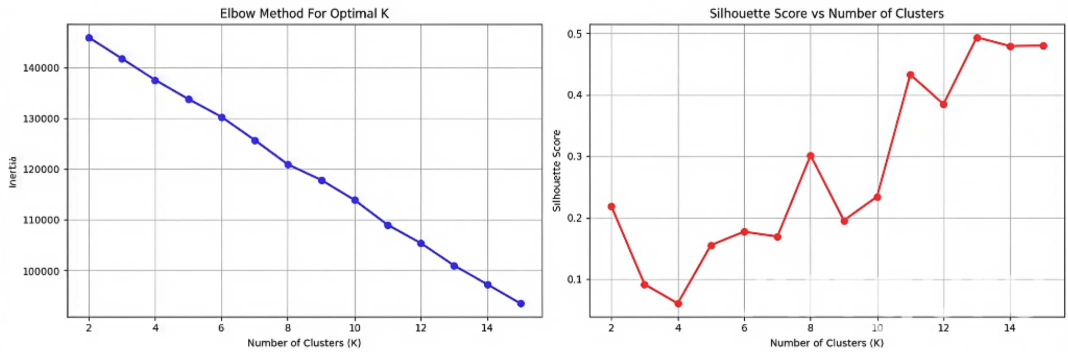


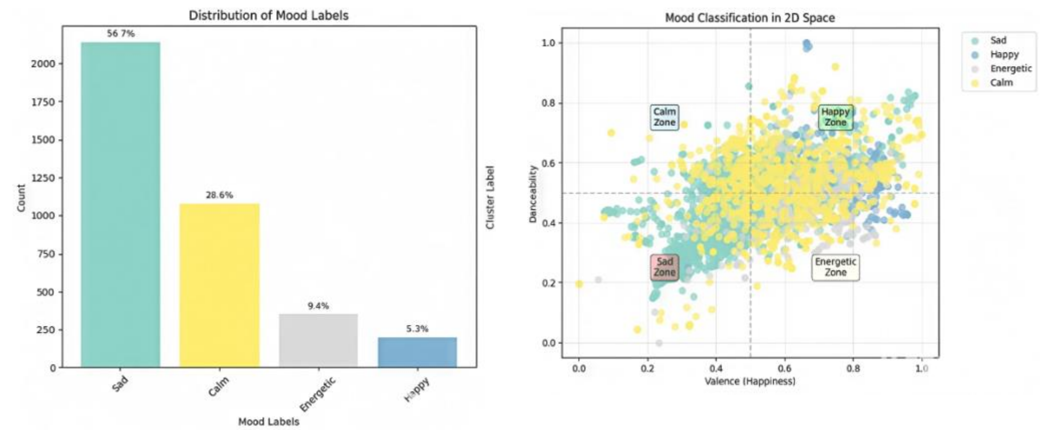
Fig. 3. Elbow Method & Silhouette Score

Each of the 15 clusters was analyzed based on the centroid values of valence and danceability. To make these clusters effective, we made a rule-based mapping, which was then assigned mood labels according to thresholds derived (see Table 1). Valence and danceability are representative of emotional characterization used in music recommendation that we made, where valence refers to the degree of positive or negative emotion in a track of music, while danceability demonstrates rhythmic suitability for movement [45].

Table 1. Emotion Classification [45]

Mood	**Valence**	**Danceability**
Happy	High (≥ 0.65)	High (≥ 0.5)
Energetic	High (≥ 0.6)	Low (< 0.5)
Calm	Moderate (> 0.5)	Moderate (~0.5)
Sad	Low (< 0.5)	Low (< 0.5)

Based on this classification (Table 1), the distribution of users across the four primary mood categories is presented in Fig. 4(a). To visualize how users were positioned emotionally, a 2D scatter plot was generated in valence–danceability space, as shown in Fig. 4(b). Each point represents a user cluster, color-coded by mood label. Zones were defined according to the classification thresholds. This clustering step provided an interpretable emotional signature for each user, later used as a categorical feature in the recommendation model. By embedding emotional preferences into the system, the recommendation process becomes more personalized and effectively relevant.



(a) (b)

Fig. 4. (a) Mood Label Distribution and (b) 2D Mood Space (Valence vs Danceability)

The model was configured using the WARP-kos (Weighted Approximate-Rank Pairwise with Kernelized Optimization Sampling) loss function, which focuses on optimizing ranking performance by learning from pairwise comparisons between positive and sampled negative items. This loss is well-suited for top-N recommendation tasks and improves learning efficiency, especially in large and sparse datasets. The training inputs consisted of a binary user-item interaction matrix, indicating whether a user had engaged with a track, and a sparse item feature matrix, containing emotional bins (e.g., valence: high, tempo: medium) and genre indicators. Through 1,200 training epochs, the model successfully learned latent representations that encode both personalized behavioral patterns and emotionally aware content preferences, enabling it to generate music recommendations that align with both long-term and situational user moods. The evaluation was performed using the Precision and Recall, with the final results shown in Table 2 below:

Table 2. Evaluation Results

Metric	**Score**
Precision@10	0.6209
Recall@10	0.4663

These results indicate that the system retrieved, on average, 6 out of 10 relevant items per user, while also capturing nearly 47% of all relevant items. This demonstrates a strong balance between ranking accuracy and coverage, validating that the effectiveness of the hybrid LightFM model was enhanced with emotional and behavioral signals. To contextualize these results, Table 3 compares the proposed method with benchmarks for similar or hybrid recommendations using LightFM and emotion-aware approaches.

Table 3. Accuracy Comparison

Reference	Approach	Result
Proposed Methods	Emotion-enhanced hybrid LightFM (Wrao-kos) with valence/danceability	Precision@10 0.6209, Recall@10 0.4663
Standard LightFM with WARP[23]	Hybrid LightFM with WRAP	Precision@10 0.10-0.36, Recall@10 0.06-0.39
Emotion-driven FCRA Hybrid[46]	Deep learning (full convolutional recurrent attention) with emotional features	Precision@10 0.885-0.902, Recall@10 0.815-0.832
Other emotion-aware systems[10]	Various (CNN, Deep Bayesian)	NDCG > 0.70

In this study, the proposed method results are notably strong for the LightFM-based approach in implicit-feedback music recommendation, outperforming many standard WARP implementations while incorporating an interpretable emotional signal. Higher scores in recent deep emotion-aware models highlight opportunities for further integration of advanced neural techniques.

Despite the proposed system achieving strong performance (Precision@10: 0.6209; Recall@10: 0.4663), limitations persisted. The inherent sparsity of the user-item interaction matrix, arising from a large number of tracks and relatively few interactions per user, made it difficult to capture long-tail preferences or nuanced listening behaviors, thereby reducing the model's ability to generalize preferences for items with limited interactions primary challenge in music recommendation systems. Additionally, the ambiguity in implicit user signals, such as play duration and skip behavior, affected the precision of learned preferences, as long listening durations might reflect passive consumption rather than genuine enjoyment, while skips could be context-dependent (e.g., mood, environment) rather than a true dislike. The rule-based clustering to assign mood labels (e.g., "Happy", "Sad") based on valence and danceability, through an interpretable approach, oversimplifies the subjective and multifaceted nature of musical emotion, with some tracks failing between categories or evoking varying feelings depending on the listener's context.

Furthermore, despite the hybrid model's use of item features to mitigate the cold-start problem, it remained vulnerable to new users with minimal history or tracks with atypical, underrepresented feature combinations, often leading to generic or popular recommendations that diminished personalization. Finally, the reliance on static, binned audio features constrained the model's adaptability to short-term emotional shifts or contextual needs, critical in an affective recommendation system, as it could not accommodate dynamic preference changes over time or variations on factors like time of day or activity.

6. Conclusion

This paper presents an emotion-aware music recommendation system that we developed using the LightFM hybrid matrix factorization framework, explicitly integrating implicit user feedback with audio-derived emotional features through mood-based user clustering. We utilize listening behavior signals such as play duration and skip patterns alongside affective audio attributes, including valence and danceability, to construct emotionally informed user and item representations. By embedding these representations within a hybrid collaborative–content-based model, we enable the system to generate personalized music recommendations that align not only with historical preferences but also with users' underlying emotional tendencies. The experimental evaluation conducted on real-world datasets demonstrates the effectiveness of this approach, achieving a Precision@10 of 0.6209 and a Recall@10 of 0.4663. These results consistently outperform traditional collaborative filtering and content-based baselines, particularly in sparse feedback settings, thereby confirming that emotional context significantly enhances recommendation quality when combined with implicit behavioral data.

Despite these promising results, this study identifies several limitations that motivate further research. We observe that data sparsity remains a challenge, particularly for capturing long-tail user preferences that appear infrequently in interaction logs. In addition, implicit feedback signals, while scalable, introduce ambiguity because listening behavior does not always reflect true user satisfaction. The mood-based clustering approach that we adopt relies on rule-based interpretations of audio features, which may oversimplify the subjective and dynamic nature of human emotions. Although the inclusion of content features partially mitigates the cold-start problem, new users and newly released tracks still suffer from limited personalization. Moreover, the use of static user and item features constrains the model's ability to adapt to temporal changes, situational contexts, or evolving emotional states over time.

Overall, this research advances the field of personalized music recommendation by demonstrating how emotional and behavioral dimensions can be effectively integrated within a lightweight hybrid recommendation framework. We show that combining matrix factorization with emotion-aware user modeling leads to a more intuitive and responsive recommendation experience without excessive computational complexity. Building on this foundation, future work may address the identified limitations by incorporating temporal and contextual signals, adopting sequential models such as recurrent neural networks or Transformers, and leveraging multimodal emotion data from text, images, or physiological signals. Additionally, we encourage the exploration of meta-learning and adaptive learning strategies to improve cold-start performance and long-term personalization.

References

[1] H. J. Chong, H. J. Kim, and B. Kim, "Scoping Review on the Use of Music for Emotion Regulation," *Behavioral Sciences*, vol. 14, no. 9, p. 793, Sept. 2024, doi: 10.3390/bs14090793.

[2] W.-J. He, W.-C. Wong, and A. N.-N. Hui, "Emotional Reactions Mediate the Effect of Music Listening on Creative Thinking: Perspective of the Arousal-and-Mood Hypothesis," *Front. Psychol.*, vol. 8, p. 1680, Sept. 2017, doi: 10.3389/fpsyg.2017.01680.

[3] H. Wu, D. Wang, and L. Zhou, "Tunes that move us: the impact of music-induced emotions on prosocial decision-making," *Front. Psychol.*, vol. 15, p. 1453808, Jan. 2025, doi: 10.3389/fpsyg.2024.1453808.

[4] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, June 2005, doi: 10.1109/TKDE.2005.99.

[5] Ò. Celma, *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-13287-2.

[6] B. Han, S. Rho, S. Jun, and E. Hwang, "Music emotion classification and context-based music recommendation," *Multimed Tools Appl*, vol. 47, no. 3, pp. 433–460, May 2010, doi: 10.1007/s11042-009-0332-6.

[7] C.-Y. Wang, Y.-C. Wang, and S.-C. T. Chou, "A Context and Emotion Aware System for Personalized Music Recommendation," 2018.

[8] D. Rozhevskii, J. Zhu, and B. Zhao, "Psychologically-Inspired Music Recommendation System," May 06, 2022, *arXiv*: arXiv:2205.03459. doi: 10.48550/arXiv.2205.03459.

[9] H. Tran, T. Le, A. Do, T. Vu, S. Bogaerts, and B. Howard, "Emotion-Aware Music Recommendation," *AAAI*, vol. 37, no. 13, pp. 16087–16095, June 2023, doi: 10.1609/aaai.v37i13.26911.

[10] E. Jing *et al.*, "Emotion-aware Personalized Music Recommendation with a Heterogeneity-aware Deep Bayesian Network," Nov. 29, 2024, *arXiv*: arXiv:2406.14090. doi: 10.48550/arXiv.2406.14090.

[11] T. Babu, R. R. Nair, and G. A, “Emotion-Aware Music Recommendation System: Enhancing User Experience Through Real-Time Emotional Context,” Nov. 17, 2023, *arXiv*: arXiv:2311.10796. doi: 10.48550/arXiv.2311.10796.

[12] H. Lee *et al.*, “GlobalMood: A cross-cultural benchmark for music emotion recognition,” June 28, 2025, *arXiv*: arXiv:2505.09539. doi: 10.48550/arXiv.2505.09539.

[13] A. Chavali and R. Menezes, “Words to Waves: Emotion-Adaptive Music Recommendation System,” Sept. 17, 2025, *arXiv*: arXiv:2510.21724. doi: 10.48550/arXiv.2510.21724.

[14] S. Wang, C. Xu, A. S. Ding, and Z. Tang, “A Novel Emotion-Aware Hybrid Music Recommendation Method Using Deep Neural Network,” *Electronics*, vol. 10, no. 15, p. 1769, July 2021, doi: 10.3390/electronics10151769.

[15] Y. Tian, “Music emotion representation based on non-negative matrix factorization algorithm and user label information,” *PeerJ Computer Science*, vol. 9, p. e1590, Sept. 2023, doi: 10.7717/peerj-cs.1590.

[16] S. Yousefian Jazi, M. Kaedi, and A. Fatemi, “An emotion-aware music recommender system: bridging the user’s interaction and music recommendation,” *Multimed Tools Appl*, vol. 80, no. 9, pp. 13559–13574, Apr. 2021, doi: 10.1007/s11042-020-10386-7.

[17] M. Quadrana, P. Cremonesi, and D. Jannach, “Sequence-Aware Recommender Systems,” *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–36, July 2019, doi: 10.1145/3190616.

[18] M. Schedl, E. Gómez, and J. Urbano, “Music Information Retrieval: Recent Developments and Applications,” *FNT in Information Retrieval*, vol. 8, no. 2–3, pp. 127–261, 2014, doi: 10.1561/1500000042.

[19] M. Schedl, P. Knees, B. McFee, D. Bogdanov, and M. Kaminskas, “Music Recommender Systems,” in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds., Boston, MA: Springer US, 2015, pp. 453–492. doi: 10.1007/978-1-4899-7637-6_13.

[20] Z. Liu, W. Xu, W. Zhang, and Q. Jiang, “An emotion-based personalized music recommendation framework for emotion improvement,” *Information Processing & Management*, vol. 60, no. 3, p. 103256, May 2023, doi: 10.1016/j.ipm.2022.103256.

[21] H. Koneru, S. Dargan, and S. Nagpal, “An Implicit Feedback Music Recommender System”.

[22] R. G. Pawar, S. U. Ghumbre, R. R. Deshmukh, and K. R. Kolhe, “A Hybrid Approach towards Improving Performance of Recommender System Using Matrix Factorization Techniques,” *International Journal of Future Generation Communication and Networking*, vol. 13, no. 4, 2020,

[23] M. Kula, “Metadata Embeddings for User and Item Cold-start Recommendations,” July 30, 2015, *arXiv*: arXiv:1507.08439. doi: 10.48550/arXiv.1507.08439.

[24] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, July 2002, doi: 10.1109/TSA.2002.800560.

[25] F. Gouyon and S. Dixon, “A Review of Automatic Rhythm Description Systems,” *Computer Music Journal*, vol. 29, no. 1, pp. 34–54, Mar. 2005, doi: 10.1162/comj.2005.29.1.34.

[26] E. Pampalk, “Computational Models of Music Similarity and their Application in Music Information Retrieval,” Doctoral dissertation, Vienna University of Technology, 2006. [Online]. Available: https://www.ofai.at/~elias.pampalk/publications/pampalk06thesis.pdf

[27] P. Knees and M. Schedl, “A survey of music similarity and recommendation from music context data,” *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 10, no. 1, pp. 1–21, Dec. 2013, doi: 10.1145/2542205.2542206.

[28] R. Mayer, R. Neumayer, and A. Rauber, “Combination of audio and lyrics features for genre classification in digital audio collections,” in *Proceedings of the 16th ACM international conference on Multimedia*, Vancouver British Columbia Canada: ACM, Oct. 2008, pp. 159–168. doi: 10.1145/1459359.1459382.

[29] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, “Columbia University LabROSA, EE Dept.”.

[30] K. Seyerlehner, G. Widmer, and P. Knees, “A Comparison of Human, Automatic and Collaborative Music Genre Classification and User Centric Evaluation of Genre Classification Systems,” in *Adaptive Multimedia Retrieval. Context, Exploration, and Fusion*, vol. 6817, M. Detyniecki, P. Knees, A. Nürnberger, M. Schedl, and S. Stober, Eds., in Lecture Notes in Computer Science, vol. 6817. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 118–131. doi: 10.1007/978-3-642-27169-4_9.

[31] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in *Proceedings of the 1994 ACM conference on Computer supported cooperative work - CSCW ’94*, Chapel Hill, North Carolina, United States: ACM Press, 1994, pp. 175–186. doi: 10.1145/192844.192905.

[32] U. Shardanand and P. Maes, “Social information filtering: algorithms for automating ‘word of mouth,’” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’95*, Denver, Colorado, United States: ACM Press, 1995, pp. 210–217.

[33] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in *Proceedings of the 10th international conference on World Wide Web*, Hong Kong Hong Kong: ACM, Apr. 2001, pp. 285–295. doi: 10.1145/371920.372071.

[34] Y. Koren, R. Bell, and C. Volinsky, “Matrix Factorization Techniques for Recommender Systems,” *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009, doi: 10.1109/MC.2009.263.

[35] Y. Koren, “Factorization meets the neighborhood: a multifaceted collaborative filtering model,” in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, Las Vegas Nevada USA: ACM, Aug. 2008, pp. 426–434. doi: 10.1145/1401890.1401944.

[36] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, “Methods and metrics for cold-start recommendations,” in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, Tampere Finland: ACM, Aug. 2002, pp. 253–260. doi: 10.1145/564376.564421.

[37] R. Burke, “Hybrid Recommender Systems: Survey and Experiments,” *User Model User-Adap Inter*, vol. 12, no. 4, pp. 331–370, Nov. 2002, doi: 10.1023/A:1021240730564.

[38] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural Collaborative Filtering,” in *Proceedings of the 26th International Conference on World Wide Web*, Perth Australia: International World Wide Web Conferences Steering Committee, Apr. 2017, pp. 173–182.

[39] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, “Graph Neural Networks in Recommender Systems: A Survey,” 2020, *arXiv*. doi: 10.48550/ARXIV.2011.02260.

[40] M. Quadrana, A. Karatzoglou, B. Hidasi, and P. Cremonesi, “Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks,” in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, Aug. 2017, pp. 130–137. doi: 10.1145/3109859.3109896.

[41] D. Jannach, L. Lerche, I. Kamehkhosh, and M. Jugovac, “What recommenders recommend: an analysis of recommendation biases and possible countermeasures,” *User Model User-Adap Inter*, vol. 25, no. 5, pp. 427–491, Dec. 2015, doi: 10.1007/s11257-015-9165-3.

[42] A. Khali, “Shopify & Spotify Streaming History Dataset.” Apr. 02, 2024. [Online]. Available: https://www.kaggle.com/datasets/arshmankhalid/shopify-streaming-history-dataset

[43] A. M and van der Velde, “Spotify Playlists.” 2021. [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/spotify-playlists

[44] M. Pandya, “Spotify Tracks Dataset.” 2023. doi: https://doi.org/10.34740/kaggle/dsv/4372070.

[45] J. A. Russell, “A circumplex model of affect.,” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.

[46] L. Zhao, G. Liu, S. Yan, and J. Zhang, “Emotion-driven music recommendation system based on fully convolutional recurrent attention networks and collaborative filtering,” *Alexandria Engineering Journal*, vol. 125, pp. 354–366, June 2025, doi: 10.1016/j.aej.2025.03.114.