Project Report - Spring 2025

Emotion-Aware Music Recommendation System Using NLP

May 1, 2025

1 Introduction

Music has been shown in research to be an effective influence on emotional regulation, mental health, and user experience. This in turn strengthens the case for systems that can interpret and respond on emotional cues. The goal of this project is to create a system that recommends music based on the user's emotional state which can be manually inputted by the user to give the machine insight into his or her thoughts. Our system will fine tune a pre-trained language model to recognize emotions in textual data and map them to a database of songs tagged with relevant emotional qualities which in turn will be stored in a vector db.

1.1 Problem Statement & Motivation:

People typically seek music that either compliments or contradicts their mood. However, a direct relationship between user emotion and music recommendation is not straightforward and shouldn't be the case. We hope to bridge this gap by leveraging current advancements in Natural Language Processing (NLP) and sentiment/emotion analysis to provide tailored song recommendations based on context. Our motivations:

1. **Personalized Experience:** Enabling users to discover music that aligns with their feelings in real time.

2. **Novelty:** Existing recommendation systems often rely on collaborative filtering or genre preferences; incorporating explicit emotional cues from text can improve accuracy and user satisfaction.

2 Related Work and Literature Review

Past music recommendation research has primarily relied on content-based and collaborative filtering. For example, Spotify and Pandora make music recommendations based on the user's listening history and audio attributes. However, they are unaware of the consumers' emotional state at any particular time.

In affect-sensitive music recommendation, approaches like MoodPlay and Emotify have explored the mapping of emotions to music via tags or user ratings. Similarly, Schedl et al. (2018) [1] studies the role of mood in music recommendation systems as citing the effectiveness of context-aware methods. Their work emphasizes how important psychological models of emotion are in personaling music suggestions. Obviously, mood-aware recommendations lead to better satisfaction compared to standard approaches. Emotion should be a primary factor, and it is the motivation for our project in evaluating user's emotions to pick the best song.

In text-based emotion recognition, Demszky et al.'s (2020) [2] paper shows the possibility of fine-tuning transformer models for nuanced emotion classification, which we aim to build upon. It had a manually annotated corpus of 58k reddit comments with 27 distinct emotion labels. It showed that BERT models outperformed traditional machine learning approaches for nuanced classification. That's why we are adopting a similar model architecture to replicate their performance in the context of music recommendation.

Delbouy's et al. [3] propose music2vec to create a neural embedding model for music that captures semantic similarity between songs. This approach motivated us to use emotion vectors to represent songs. It also motivated us to use cosine similarity to compare song emotion vectors and compare them to text profiles using cosine similarity.

Yang and Chen's paper [4] explored ML techniques for mood detection in music. They evaluated SVMs, decision trees, and other models for classifying mood. This shows that combining lyrics and audio data improved performance over single modality models. This shows that to further improve our models in a full-on research project, we should include audio as well.

3 Methodology

3.1 Approach

1. **Text Emotion Extraction:** Train a transformer-based model (distilbert-base-uncased) to forecast one or more emotion labels (sadness, joy, anger, etc.) from the text of users.

2. **Music Emotion Representation:** Each song in our database will also have an emotion vector either derived from manual tags, previous research datasets, or through automatic lyric analysis.

3. **Matching and Recommendation:** Compute similarity between the user emotion vector and the song emotion vectors (e.g., via cosine similarity) to recommend the top k songs.

3.2 Datasets

We plan to use two types of datasets in this project:

- **Textual Emotion Dataset:** We will be using the GoEmotions dataset [**?**]—a large-scale dataset of Reddit comments that are hand-annotated for 27 emotions. This dataset will serve as the basis for fine-tuning our text emotion classification model.

- **Music Dataset:** For music, we intend to utilize a pre-existing dataset where songs are labeled with emotion or mood. We did not end up finding a suitable, publicly available dataset with these requirements, so we ended up using the Million Song

Dataset (without explicit emotion labels). So that we can still continue, what we did was that we took lyrics from the dataset, and passed it through our now, fine-tuned transformer on the GoEmotions dataset. This resulted in an emotion probability vector for each song's lyrics that were L2 normalized and then subsequently stored in FAISS index which is Facebook AI Similarity Search which can let us have an easy look-up as each vector is associated with each song's metadata.

3.3 Algorithms

We will use:

- **Transformer-based model** Distilbert-base-uncased model fine-tuned for multi-label emotion classification in text.
- **Lyric or metadata-based model** to label the emotional labels of the songs (if not already labeled).
- **Similarity measure** Cosine similarity to match user emotion vectors against the song database.

4 Evaluation

4.1 Metrics

- **Recommendation:** LLM-evaluated accuracy based on baseline and model outputs.
- **User Satisfaction Study:** User-evaluated accuracy based on baseline and model outputs.

4.2 Baseline Comparison

To evaluate the effectiveness of our system, we compare it against one baseline:

- **Lyric similarity:** Suggest songs solely based on user query similarity to song lyrics.

By contrasting with these baselines, we can tease out the benefit of using fine-tuned transformer models for emotion detection and continuous-valued emotion vectors for recommendation.

5 Experimentation and Results

We followed our proposal and created a transformer-based model to forecast one of more emotion labels from the text. We also used the same model to evaluate emotions of the songs from the million dollar dataset. We then indexed each song by its emotion, and returned the top-k songs for a person's emotion.

We evaluated the success of our results by comparing against the baseline, which simply compared song lyrics to user queries using cosine similarity. We used ChatGPT, Claude, and Gemini to evaluate the percentage increase in recommendation quality. We created 10 prompts, and evaluated the results using the llms.

Here are the results:

Model	**Improvement over Baseline (%)**
ChatGPT	53%
Claude	35%
Gemini	68%

Table 1: Percentage improvement in recommendation quality over the baseline, as judged by different LLMs.

As you can see, our model outperformed the baseline by an average of 52%. This shows that the emotion based BERT model is successful in making music recommendations.

Additionally, we conducted user testing to gauge the improvement. After conducting a study by polling random people off the street, we asked them to say how many out of the 5 songs recommended identified with their current emotional state/user input. On average, the BERT model received 3.47/5 while the baseline model received a score of 1.84/5.

Based on both user and llm generated results, our model provides a great improvement in accuracy.

6 Group Members

Roshan Arokiaraj
Subhajit Das
Sarvesh Sathish

References

[1] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7(2):95–116, 2018.

[2] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 4040–4054. Association for Computational Linguistics, 2020.

[3] Romain Delbouys, Romain Hennequin, Daniel Picard, Frank Nielsen, and Frédéric Bimbot. Music2vec: Symbolic music embedding based on skip-gram. In *Proceedings of*

the 19th International Society for Music Information Retrieval Conference (ISMIR), pages 716–722, 2018.

[4] Yi-Hsuan Yang and Homer H Chen. Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):40:1–40:30, 2012.