Project Title: Context Aware Music Recommendation System

Team members:

Full name	ASU ID
Spoorthi Uday Karakaraddi	1229585476
Tanmai Mukku	1229609006
Sai Siddharth Vemuri	1229566925
Pavan Kalyan Reddy Cherupally	1229610085

Problem statement:

The problem under investigation centers on the limitations of current music recommendation systems, which primarily leverage user preferences and song metadata for suggesting tracks. These systems often fail to consider the user's current context, such as their emotional state, physical environment, and temporal factors, which can greatly influence their music choices. The lack of contextual understanding can lead to suboptimal recommendations that do not align with the user's present needs or desires. This project seeks to enhance the user experience by incorporating contextual data into our recommendation algorithm, using data mining techniques to deliver more precise and contextually relevant music suggestions.

Related work:

[1] Pichl, M., Zangerle, E., and Specht, G. 2015. Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name? *In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 2015, 1360–1365*. https://doi.org/10.1109/ICDMW.2015.145

[2] Baxter, M., Ha, L., Perfiliev, K., and Sayre, N. 2021. Context-Based Music Recommendation Algorithm Evaluation. arXiv:2112.10612. Retrieved from DOI: https://doi.org/10.48550/arXiv.2112.10612

[3] Hamed Zamani, Markus Schedl, Paul Lamere, and Ching-Wei Chen. 2019. An Analysis of Approaches Taken in the ACM RecSys Challenge 2018 for Automatic Music Playlist Continuation. ACM Trans. Intell. Syst. Technol. 10, 5, Article 57 (September 2019), 21 pages. https://doi.org/10.1145/3344257

Initial hypothesis:

The primary research question is: Can the incorporation of contextual information such as user mood significantly improve the relevance of music recommendations? We hypothesize that integrating such contextual elements will result in a marked improvement in user satisfaction with the recommendations provided.

Dataset(s):
We will be using 2 publicly available datasets

Dataset 1

Dataset source (link and reference)	Poddar, A., Zangerle, E., and Yang, Y.-H. 2020. #nowplaying-rs: A New Benchmark Dataset for Building Context-Aware Music Recommender Systems. Zenodo. Available at: https://zenodo.org/records/3248543
Number of instances	• 11.6 million music listening events (LEs) of 139K users • 346K tracks from Twitter
Number of features	20
Class distribution (# instances in each class, if applicable)	Not Applicable
Dataset splits	Training Set: 60% Validation Set: 20% Testing Set: 20%
Preprocessing steps	- Handling Missing Values - Data Cleaning - Feature Selection - Feature Scaling - DateTime Parsing - Exploratory Data Analysis - Train-Validate-Test Split

Dataset 2

Dataset source (link and reference)	Spotify. 2023. The Million Playlist Dataset... Remastered. Spotify Research. Available at: https://research.atspotify.com/2020/09/the-million-playlist-dataset-remastered/
Number of instances	• 1 million playlists • 2 million unique tracks • 300,000 artists
Number of features	18
Class distribution (# instances in each class, if applicable)	Not Applicable
Dataset splits	Training Set: 60% Validation Set: 20% Testing Set: 20%
Preprocessing steps	- Handling Missing Values - Data Cleaning - Feature Selection - Feature Scaling - DateTime Parsing - Exploratory Data Analysis - Train-Test Split

In addition to this we will be using Spotify's API [https://developer.spotify.com/documentation/web-api] for fetching the track features using track Ids to augment the dataset, if required.

Method(s):

The approach will use a hybrid algorithm that combines collaborative filtering with context-aware techniques to create personalized music recommendations. The novelty lies in the multifaceted context-modeling that extends beyond playlist names to include contextual factors. Implementation will rely on Python libraries such as scikit-learn for machine learning, pandas for data manipulation, and possibly TensorFlow or PyTorch if deep learning techniques are applied for collaborative filtering. Compared to the state of the art, this method aims to provide a more holistic view of the user's situation, enhancing the personalization of recommendations. Specifically, incorporating the principles of Factorization Machines for handling sparse data and enabling interaction among features (users, items, context) efficiently, this method stands to significantly improve recommendation relevance by considering user mood as a contextual factor.

Evaluation:

Quantitatively the performance of our recommendation system can be measured using evaluation metrics such as accuracy, precision, recall, and F1 score calculated on the test datasets. Ranking metrics such as Mean Average Precision and Normalized Discounted Cumulative Gain, and utility-based metrics such as Expected Utility along with user satisfaction surveys or qualitative feedback from users can provide insight into the quality of recommendations. Comparing the proposed method(s) to existing methods or baseline approaches can be done by conducting A/B testing or using offline evaluation techniques like holdout validation or cross-validation.

Management plan:

We plan to divide the work based on individual expertise and interest, ensuring that each team member contributes effectively. Accountability will be upheld through regular progress updates, milestone tracking, and peer reviews. Discord will serve as our primary communication platform for weekly sync-ups, while a shared project repository on GitHub will facilitate code collaboration and version control. Additionally, we will maintain a detailed project timeline on GitHub Projects to track task completion and ensure timely progress.