

Team 2 Database Project Proposal

Our team plans to design a recommendation algorithm with machine learning, and graph modeling of the spotify database using Neo4j. We plan to offer a different style of music recommendation algorithm that will suggest more unusual or atypical music choices based on a user's predefined preferences. This will allow the users to explore different genres and types of music that are similar in attributes with a unique exploratory matching.

SMART:

1. **Specific:** We plan to design alternative, music recommendation algorithms compared to typical spotify music recommendations. We will showcase this music recommendation with machine learning algorithms.
2. **Measurable:** The true success of our algorithms would be achieved through the spotify users. But in lieu of these users, our success is based on the robustness of our algorithm. Another measurable result is that Neo4j will create a graph database mapping the relationships between users, songs, genres, and other musical attributes.
3. **Achievable:** Spotify's API will allow us access to metadata about musical choices such as genre, bpm, artist, including user listening history, preferences, genre information, and song attributes etc. So the data will need cleaning like all datasets, but most of the key variables will be available which provides many variables to manipulate in our recommendation system.
4. **Relevant:** There are already recommendation algorithms for music or other media, however our algorithm promotes different pairing of music suggestions. These unusual recommendations offer a different experience of recommendation which some users would enjoy.
5. **Time-oriented:** We aim to complete this project within the semester, which means our open-source algorithm will be robust enough to demonstrate our music recommendation system.

Neo4j and Machine Learning Roles:

Neo4j will graph our database and also manage the database. The developed nodes in our graph showcase the natural relationships in a complex data structure. We also plan to use the pre-trained algorithm to train our dataset. The algorithm will be open-source and can be tweaked accordingly. The machine learning algorithm will help us in identifying the underlying patterns in the dataset and will be the backbone of our recommendation system. There are multiple algorithms which we plan to implement such as Collaborative Filtering and Matrix Factorisation. Both of these algorithms the contents are filtered by user-based and item-based.

DATA SOURCE:

<https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset?resource=download>

Team Github: <https://github.com/sdf5y/DataWarehousing>