Mood Detection with Music Recommendation System

Sonam Rinjin Sherpa
Masters of Science in Data Science
Anderson College of Business and Computing
Regis University, Denver, CO, USA
srinjinsherpa@regis.edu

CONTENTS

I. Executive Summary

The rapid expansion of digital music streaming platforms has transformed the way people discover and consume music [1]. Modern streaming services provide users with access to millions of songs in diverse genres, artists, and cultural contexts. Although this availability of music has significantly improved accessibility and user choice, it has also introduced challenges related to information overload. With such large catalogs of available content, users often face difficulty identifying songs that align with their preferences, listening context, or emotional state. As a result, recommendation systems have become essential components of modern music streaming platforms, helping users navigate large music collections and discover relevant content efficiently [2].

This project develops a mood-sensitive music recommendation system using publicly available Spotify song datasets. The objective of the project is to explore how measurable musical attributes can be used to approximate the emotional characteristics of songs and generate meaningful recommendations. The datasets used in this study contain song-level metadata and numerical audio features extracted from Spotify's audio analysis algorithms. These features include attributes such as danceability, energy, tempo, loudness, acousticness, and valence, which collectively describe the structural and perceptual properties of music.

The recommendation system implemented in this project follows a content-based filtering approach. In this system, each song is represented as a vector composed of selected audio features. The similarity between songs is calculated using feature-based similarity measures, allowing the system to identify tracks with comparable acoustic properties. By computing similarity scores between songs, the system generates Top-N recommendations for a selected input track or mood preference.

The project workflow includes data cleaning, feature selection, normalization, and exploratory data analysis to examine the relationships between musical attributes. A similarity-based recommendation model was developed using cosine similarity. The results demonstrate that the audio features extracted from Spotify provide meaningful signals to present musical structure and emotional tone. In general, the system illustrates how interpretable data science techniques can be applied to develop practical music recommendation systems that support mood-based music discovery.

II. Introduction/Background

Due to this rise in digital music streaming, it becomes difficult to find a suitable song that is in line with your mood, situation, and preferencex. Hence, it is not surprising that a recommendation system is a vital component of modern music streaming systems.

From a data science point of view, a recommendation system is a vital example of a real-world application of machine learning techniques [3]. It involves a series of processes such as data preprocessing, feature engineering, computing the similarity between songs, and evaluating performance, among others. By examining patterns in data, a recommendation system helps users explore relevant songs in a massive collection of songs.

According to [4], mood plays a critical role in music selection. People choose music according to their mood and situation. For example, people may listen to relaxing music when studying, while they listen to energetic music when doing exercises.

However, the area of music information retrieval indicates that the audio features of songs can provide information regarding the structure of the music and the tone of the song [5]. Various audio features are computed for each song using the Spotify platform. Some of the audio features include danceability, energy, loudness, tempo, acousticness, and valence. These audio features provide a numeric representation of the features of the song along with the tone of the song. A mood-aware recommendation system is developed using the audio features of Spotify datasets. It is possible to find similarities between songs using their numerical features.

In this project, a recommendation is developed using a content-based filtering approach, rather than the use of the deep learning method. Content-based filtering is a method in which the features of the song are considered to recommend similar songs.

III. Problem Statement

This rapid growth of digital music streaming services has given users unprecedented levels of access to music content. Services like Spotify offer users millions of tracks across a variety of genres, artists, and cultural settings. Though this opens doors for users to enjoy a vast amount of music content, it also creates a considerable challenge for users to effectively find the music they need or want to listen to. Hence, the ability to effectively find the most relevant music content has become a major challenge for music streaming services.

Recommendation systems have become an integral part of addressing the music discovery challenge. Recommendation systems analyze user behavior, music content, and other relevant information to help users effectively find music content they need or want to listen to. There are a variety of music recommendation techniques already implemented by music streaming services. Many of these techniques rely on collaborative filtering to generate music recommendation results. Collaborative filtering techniques have proven to be effective for many recommendation systems [6]. Despite the effectiveness of collaborative filtering techniques for music recommendation, they often fail to take into account the intrinsic properties of the music content.

One of the major limitations of music recommendation systems is the lack of consideration for the emotional attributes of music content [1]. Music consumption is largely an emotional activity. Users often listen to music that matches their emotional state or the music they need to listen to at a particular time. For instance, users often listen to music that helps them concentrate while studying or music that helps them exercise or music that helps them feel better or happier. Despite the fact that music recommendation systems fail to take into account the emotional attributes of music content, music consumption is a highly emotional activity.

From the point of view of data science, the problem of incorporating the concept of mood in computational models is a complex problem [7]. The reason for this is the subjective nature of emotions. Despite the difficulty in directly quantifying emotions, musical attributes extracted from audio analysis can provide quantitative measures for the characteristics of emotions. Musical attributes such as energy, valence, tempo, and danceability describe the structural or perceptual characteristics of music that have been found to relate to emotions. By analyzing the musical attributes of different songs, the relationship between the songs can be inferred.

With the above challenges in mind, the project explores the development of a music recommendation system that is aware of the concept of mood based on audio features from the Spotify audio analysis system. The goal of the system is to find songs that have similar acoustic characteristics and provide recommendations that match the mood of the user.

This project contributes to the field of music recommendation systems by showing the application of data science techniques in the development of music recommendation systems. Besides the practical contribution of the project to the field of music recommendation systems, the project shows the potential of content-based recommendation systems.

IV. Data Collection and Preparation Methodology

A. Data Description

The dataset used in the current study consists of two Spotify tracks from publicly available datasets on Kaggle. Once the datasets are integrated, the combined dataset contains 105,477 songs and 14 variables.

The dataset contains not only the attributes of the data but also the numerical features of the audio tracks. Some of the audio features are the danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, and tempo of the audio tracks.

Audio attributes are the features of the audio tracks, and each attribute measures the audio tracks on a specific feature. Danceability measures the tracks based on the ability to dance to the music, whereas the energy attribute measures the perceived music intensity. Valence measures the positive emotions contained in the audio tracks, with higher values indicating positive and uplifting audio tracks. Table I summarizes the key characteristics of the combined Spotify dataset used in this study.

TABLE I: Summary of the Combined Spotify Dataset

Attribute	Value
Total Number of Songs	105,477
Total Number of Features	14
Numerical Features (float)	9
Integer Features	1
Categorical/Text Features	4
Audio Features Used	Danceability, Energy, Loudness, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo
Dataset Source	Combined Spotify datasets (Kaggle)

B. Data Cleaning and Preprocessing

Before it is subjected to modeling and analysis, a series of preprocessing techniques were applied to ensure data quality and consistency. The first preprocessing technique applied to the data was data cleaning, which includes removing duplicate values and identifying missing values in the data set. It is important to ensure that each track is unique in the data set to avoid any kind of bias in the similarity calculation and maintain the reliability of the system.

C. Feature Selection

After data cleaning, feature selection was carried out to determine the most important features to use to represent the similarity and mood characteristics of the music. Although the dataset contains many features, the numerical audio features are the ones chosen to be the input features to the recommendation model. These features provide a structured representation of the music attributes, which could be used to measure the similarity between the songs.

D. Feature Normalization

Since the audio features are based on different numerical scales, feature normalization was performed to standardize the values. This is a process that aims to normalize the features, ensuring that each feature has an equal contribution to the results of the similarity calculation. This is important to prevent features with large numerical ranges from dominating the results of the distance calculation.

Finally, the preprocessed data were then used for exploratory data analysis and the development of the recommendation model. This is important to ensure that the data being used for the analysis is reliable, appropriate, and consistent.

V. Analysis Methods and Implementation Details

A. Exploratory Data Analysis

Before the development of the recommendation system, exploratory data analysis (EDA) was performed to achieve a better comprehension of the characteristics of the data, as well as the patterns present in the audio features. EDA is a significant part of the data science pipeline as it helps to identify the distribution of the features, identify possible anomalies, and identify the relationships among the features, which may be useful for the development of the recommendation system [**?**]. The exploratory data analysis focused mainly on the numerical audio features obtained from the dataset.

Various visualizations, such as the distributions of the features and correlations among the features, were obtained. For example, the energy level is highly correlated with the loudness level, as louder audio is more likely to be more energetic. Similarly, the valence feature helps identify the level of positivity of the audio, which may be useful for the representation of the mood feature.

The results obtained from the exploratory data analysis were useful in selecting the appropriate audio features for the recommendation system, as well as ensuring the appropriateness of the features to represent the audio.

The insights gained from the exploratory data analysis informed the selection of relevant audio features used in the recommendation model and helped confirm that these variables contain meaningful information to represent musical similarity.

B. Mood Representation Using Audio Features

A major objective of this project is to approximate the characteristics of the mood of the song using audio features. Although the emotions evoked by the audio signals are subjective, there are some audio features that are commonly associated with particular emotions. Audio features that can be used as surrogates for the characteristics. Among the audio features, energy and valence play a major role in representing the mood. Energy is the level of intensity or the activity level of the audio signal, whereas valence is the level of emotional positivity. Songs with high energy and high valence are more likely to be lively, whereas songs with low valence may reflect a melancholic tone.

In addition, audio features such as tempo and danceability play an important role in representing the mood of the song. By combining the audio features, it is possible to develop a multi-dimensional representation of the song. This representation is based on the combination of the structural features of the song along with the audio features.

The overall workflow of the recommendation system is illustrated in Figure 1.

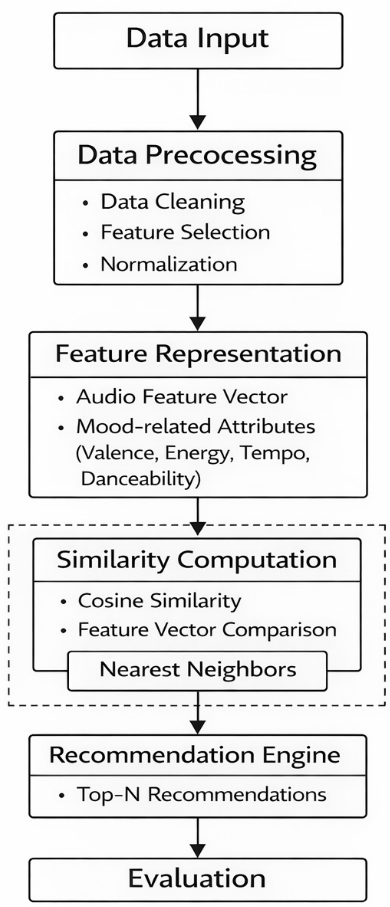

Fig. 1: System architecture for mood detection-based music recommendation system.

C. Content-Based Recommendation Model

The recommendation system proposed in this project is based on the content-based recommendation approach. A content-based recommendation system is based on the recommendation of items based on the features of the items [8]. In other words, the recommendation is based on the intrinsic characteristics of the item. In the context of the music recommendation system, the intrinsic characteristics of the item refer to the features of the audio signals. By analyzing the features of the audio signals, the system can identify the similarities among the audio signals. In this recommendation system, each audio signal is represented as a feature vector.

Content-based recommendation systems have several benefits to offer in this context. In the first place, the fact that the recommendations are based on quantifiable characteristics of the music allows the system to make recommendations without any significant user interaction data. Furthermore, the use of audio characteristics allows us to understand how and why a recommendation is made.

D. Similarity Computation (Cosine Similarity)

To find similar songs based on musical characteristics, the similarity between the song feature vectors is computed. Cosine similarity is used for this purpose. Cosine similarity is a commonly used method for recommendation systems, especially for information retrieval. This is because the cosine similarity method calculates the similarity between two vectors on the basis of the angle between the two vectors, rather than the magnitude of the two vectors.In the project, the feature vector is a normalized vector represents the song. The Cosine similarity is then used to calculate the similarity score between the two songs. A high similarity score indicates that the two songs are similar, whereas a low score indicates dissimilarity.

Finally, the system retrieves the Top-N similar songs for the given input song based on the computed similarity scores. Then, the recommendations are given based on similar songs.

E. Recommendation Generation Using Nearest Neighbours

In order to efficiently generate the recommendations, the system implemented a nearest neighbor search with the help of the normalized feature vectors. Once the similarity between the songs is computed, the system finds the nearest neighbors of the given song.

The nearest neighbor algorithm finds the songs with the nearest feature vector to the given input song by calculating the cosine similarity. These nearest neighbors to the given song share similar acoustic properties as well as mood-related attributes.

Once the nearest neighbors to the given song are found, the system finds the Top-N songs most similar to the given song as a recommendation.

VI. Results with Visualizations and Interpretation

A. Exploratory Data Analysis Results

To gain a better understanding of the structure of the audio features, exploratory data analysis was performed. This analysis sought to identify patterns in the audio features, which would play a critical role in the computation of the recommendations.

To achieve this, statistical visualizations were performed to examine the distribution of the audio features. As indicated, the audio features encompass a variety of musical structures. For example, the danceability feature represents the extent to which a particular song is danceable based on the rhythm. However, the energy feature represents the energy level of the song, among other features. Figure 2 illustrates the distribution of selected Spotify audio features across the dataset.

The results obtained from the audio features demonstrate the diversity of song types. Some of the features, such as energy, are high, indicating that the corresponding song has a high tempo, which is commonly associated with high-energy music such as electronic dance music. However, the features may demonstrate low energy, indicating a slower tempo, commonly associated with calm music such

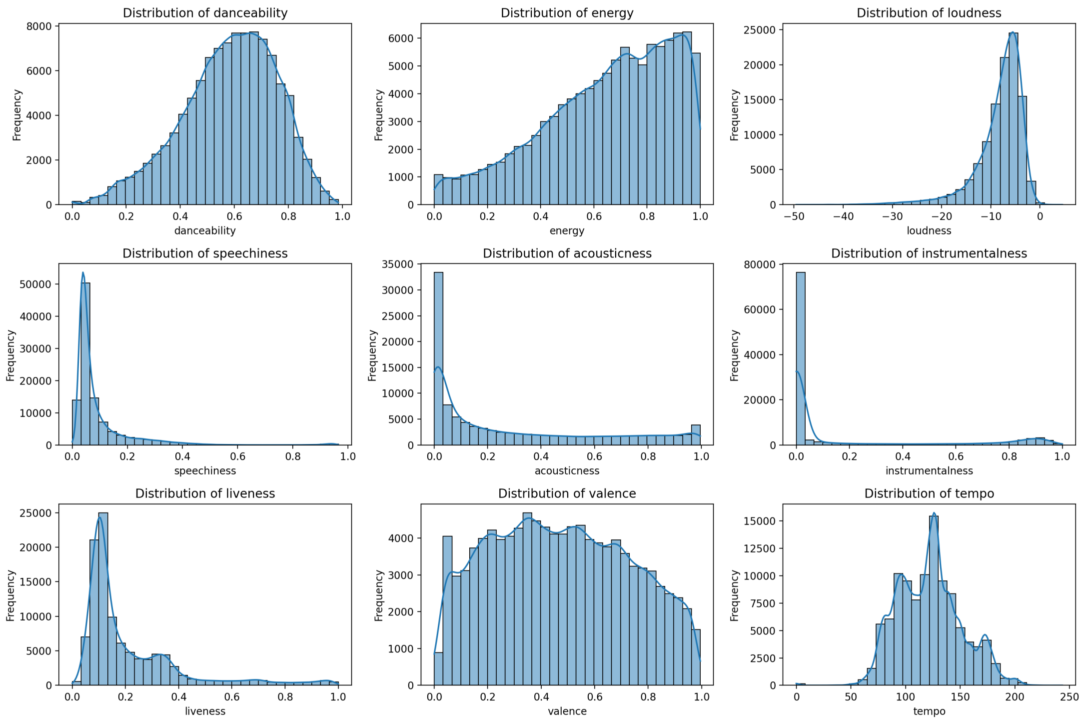

Fig. 2: Distribution of key Spotify audio features used in the recommendation system.

as acoustic or instrumental music. Figure 3 visualizes the mood distribution of songs using energy and valence features.

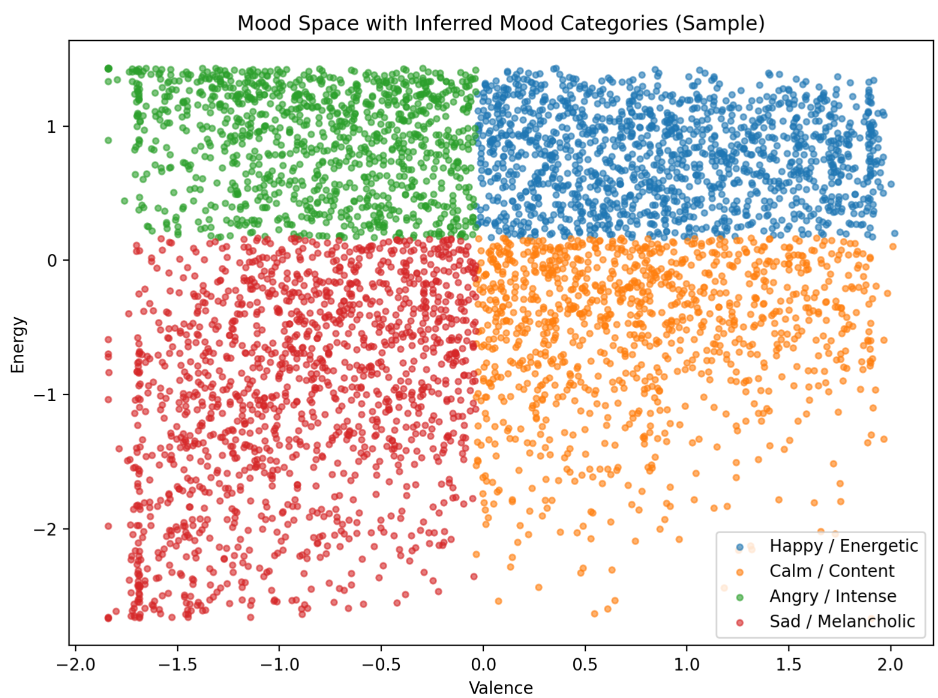

Fig. 3: Mood representation of songs using energy and valence features.

These observations confirmed that the audio features provide useful signals for representing musical similarity and mood-related characteristics within the recommendation system.

B. Recommendation System Results

The recommendation system was evaluated using a dataset of approximately 105,000 tracks from the dataset. Each track had various audio features that described the properties of the song. Once the dataset was processed, the recommendation system was implemented using a content-based filtering technique. Using this technique, every track was represented as a numerical feature vector with various audio features from the dataset. The audio features together described the structural properties of the music.

For the recommendation system to generate the output, the similarity between the tracks was computed by using the cosine similarity function and the nearest neighbors algorithm. Using the cosine similarity function, the similarity between the tracks can be computed by calculating the angular distance between the feature vectors. Using the nearest neighbors algorithm, the most similar tracks to the input track were computed.

For a given input track, it computes the similarity scores of the input track with respect to all other tracks in the dataset. Then it returns tracks with the highest similarity scores as recommendations. Essentially, these are tracks that share similar characteristics in terms of energy levels, tempo patterns, and emotional tone. Figure **??** illustrates the nearest neighbor recommendations generated for a selected query song. The recommended songs cluster closely around the query song in the feature space, indicating similarity in musical attributes.

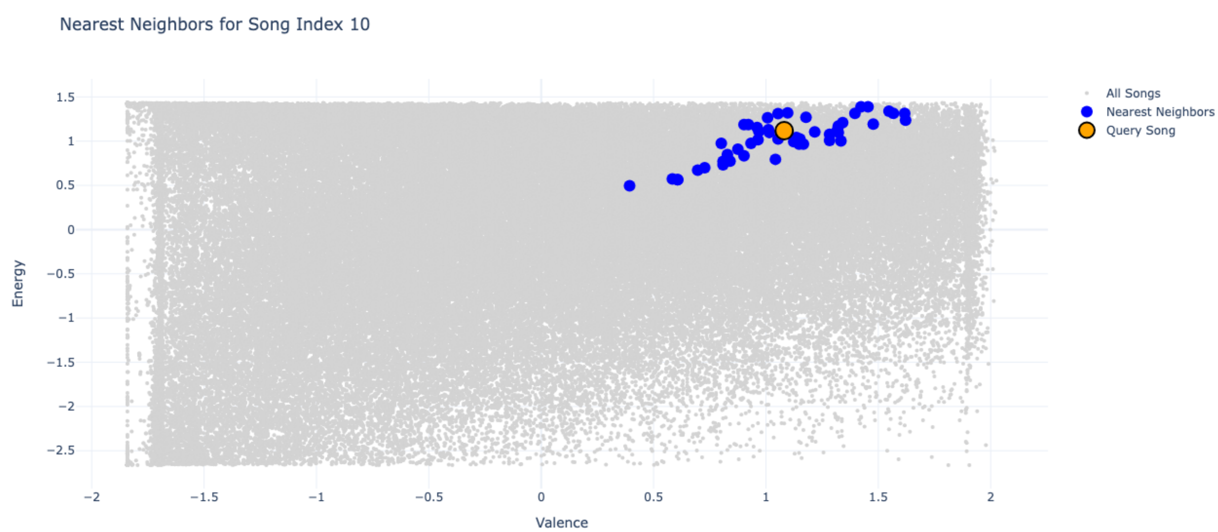

Fig. 4: Visualization of nearest neighbor recommendations for a selected query song

In effect, the developed recommendation system is highly successful in identifying tracks with similar acoustic structures and mood-related attributes. For example, tracks with high energy levels and high tempo are likely to be recommended with tracks having similar energy levels, while tracks with low energy levels and high acousticness are likely to have recommendations with similar levels of acousticness.

Therefore, using nearest neighbor retrieval, it becomes highly efficient in generating recommendations while still remaining interpretable. As a result, it becomes possible to determine why a particular track is being recommended by examining its features with respect to another track. Essentially, this is a positive attribute of a content-based recommendation system compared to more complex systems.

In conclusion, it becomes clear that the developed recommendation system proves that Spotify audio features are highly useful in determining relationships between tracks. As a result, it becomes highly successful in generating relevant music recommendations using similarity-based modeling techniques.

C. Evaluation and System Performance

Compared to conventional classification and regression tasks, there are unique challenges in evaluating a recommendation system. For example, in a number of cases, it has been seen that the performance of a recommendation system is heavily dependent on listening behavior and user preferences. As a result, it is common to see a variety of quantitative and qualitative evaluation techniques used in these systems.

In this case, the performance of the recommendation system was evaluated using a variety of techniques, including similarity-based evaluation and qualitative evaluation of the system's performance. As a result of using a content-based filtering approach in this case, it was not possible to employ a collaborative filtering approach to evaluate this system.

In this case, it is clear that the recommendation system retrieves the Top-N nearest neighbor tracks to a given input track using a variety of techniques, including cosine similarity calculations. As a result, tracks with the highest cosine similarity scores are retrieved as recommendations, where these scores are indicative of the level of similarity in terms of their acoustic features.

TABLE II: Evaluation Metrics for Recommendation System

Feature Set	**Mood Consistency**	**Avg Similarity**	**Catalog Coverage**
Extended Audio Features	0.9423	0.994	0.9927
Mood Features Only	0.9497	1.000	0.9954

From the qualitative evaluation of performance of the system, it is clear that several tracks are seen to share a variety of salient features in terms of their musical characteristics. For example, it is clear that tracks with high energy and fast tempo are seen to recommend tracks with energetic music genres such as electronic and pop music. Similarly, tracks with high levels of acousticness and lower energy are seen to recommend tracks with more tranquil music genres.

In this case, although it is not possible to determine a variety of quantitative performance metrics using this approach, it is clear that this recommendation system is competent in determining meaningful relationships between tracks using a variety of Spotify audio features. As a result, it is clear that this system is competent in performing a variety of music recommendation tasks using a variety of content-based data representations.

VII. Challenges and Limitations

Despite the successful implementation of the music recommendation system, the system development process has faced various challenges and limitations.

One of the major limitations of the system is the subjective nature of musical perception. The emotional interpretation of music is vastly different varies among listeners. In addition, the audio features used in the system provide only an approximate estimate of the mood of the music.

The content-based approach is a good method for music recommendation systems. However, the best recommendation systems today use a combination of different approaches. For example, the Spotify recommendation system is a hybrid approach that combines the content-based approach with the collaborative approach.

The use of a limited dataset might be a limitation for the project. The dataset used in the project is only part of the entire dataset available on Spotify. This is a limitation of the system. The system would have been better able to recommend diverse music.

The computational challenges of handling a huge dataset is also a potential limitation. Calculating the similarity of the audio features for thousands of songs is a computational challenge. Although the system has used the nearest neighbor approach to calculate the similarities between the audio features of the songs, the system would have to use other techniques to handle millions of songs.

Despite the limitations of the system, the project has shown that the system can provide a good foundation for the development of music recommendation systems.

VIII. Contributions and Significance

This practicum project demonstrates the application of data science methodologies to the problem of music recommendation systems. The main technical contribution of this project is the design and development of a content-based recommendation system that uses audio features to recommend songs with similar attributes.

One of the other significance of this project is the exploration of the problem of music recommendation based on moods. With the addition of features, the recommendation system provides music recommendations based on the perceived emotions of the songs.

Aside from the technical applicability, the demonstration of the potential of data science methodologies to the problem of digital music discovery. Recommendation systems play a crucial role in modern music streaming systems, and the data science methodologies demonstrated in this project provide a good starting point for further exploration and development of music recommendation systems.

IX. Future Work

The implemented recommendation system proves the effectiveness of using content-based music recommendation techniques; however, there are some opportunities to improve and extend this system in various ways.

One of these ways is to improve this system in the future by using collaborative filtering techniques in conjunction with content-based techniques. Such a system could provide more accurate recommendations using a combination of these two techniques.

Another direction in which this system could be improved in the future is to increase the data set used in this system. Using a larger data set could improve the variety of recommendations provided by this system and help it better understand the relationships between different genres of music.

Using deep learning techniques in music recommendation systems could also be a good enhancement in which this system could be improved in the future. Such techniques could provide more accurate recommendations using more complex features of music.

Using more features, such as lyrics, could also improve this system in various ways and help it provide more accurate recommendations by using more complex features of music.

X. Conclusion

This practicum presents the development of a mood-based music recommendation system that relies on audio features obtained from the Spotify platform, along with content-based filtering methods. Using the quantified audio features, the system can represent songs as feature vectors, making recommendations based on similarities between songs.

The practicum demonstrates the feasibility of integrating data preprocessing, exploratory data analysis, and similarity-based modeling to develop a functional recommendation system. In addition, the results obtained from the research prove the effectiveness of the audio features obtained from the Spotify platform, which can be utilized to identify songs with similar features.

Although the research is based on a simplified content-based recommendation system, it is effective in illustrating the underlying principles of recommendation systems. Subsequently, the research highlights the potential of data science techniques to address real-world challenges.In general, the research moves to provide a practical example of the potential of recommendation systems facilitates the exploration of songs. Furthermore, the techniques and methodologies used in this research provide the basis for the development of more sophisticated recommendation systems.

References

[1] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi, "Current challenges and visions in music recommender systems research," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, p. 95–116, Apr 2018.

[2] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*. New York, Ny: Springer Us, 2015.

[3] Y.-H. Yang and H. H. Chen, "Machine recognition of music emotion," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 3, p. 1–30, May 2012.

[4] H.-C. Chen and A. L. P. Chen, "A music recommendation system based on music and user grouping," *Journal of Intelligent Information Systems*, vol. 24, no. 2-3, p. 113–132, Mar 2005.

[5] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *9th international symposium on computer music modeling and retrieval*, vol. 4, 2012, pp. 395–410.

[6] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, "Recommender systems leveraging multimedia content," *ACM Computing Surveys*, vol. 53, no. 5, p. 1–38, Oct 2020.

[7] J. S. Lee and J. C. Lee, "Music for my mood: A music recommendation system based on context reasoning," *Lecture Notes in Computer Science*, p. 190–203, 2006.

[8] V. Moscato, A. Picariello, and G. Sperli, "An emotional recommender system for music," *IEEE Intelligent Systems*, p. 1–1, 2020.