**Music Recommendation System**

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

**CERTIFIED SPECIALIST**

**IN**

**Machine Learning and Artificial Intelligence**

submitted by

**Nandhini P S**

**Saranya Rajagopalan**

**Anisha Poulose**

**Anita John**



**ICT ACADEMY OF KERALA**

**THIRUVANANTHAPURAM, KERALA, INDIA**

**February 2023**

**List of Figures**

Figure 1: Song Recommendation System User choice……………………………13

Figure 2: List of Song Recommendations…………………………………………14

Figure 3: Spotify page of the recommended song…………………………………14

**List of Abbreviations**

1. AI - Artificial Intelligence
2. CSV - Comma Separated Value
3. MRS - Music Recommendation System
4. CBRS – Content Based Recommendation System
5. BPM – Beats Per Minute
6. ML – Machine Language
7. ICIEE - International Conference on Informatics Electrical and Electronics

**Table of Contents**

[Abstract 5](#_Toc142396737)

[1. Problem Definition 6](#_Toc142396738)

[1.1 Overview 6](#_Toc142396739)

[1.2 Problem Statement 6](#_Toc142396740)

[2. Introduction 7](#_Toc142396741)

[3. Literature Survey 9](#_Toc142396742)

[4. Data Collection 10](#_Toc142396743)

[4.1 Data Sources 10](#_Toc142396744)

[4.2 Dataset Description 10](#_Toc142396745)

[4.3 Data Preprocessing 11](#_Toc142396746)

[5. Methodology 12](#_Toc142396747)

[5.1 Methodology by Steps 12](#_Toc142396748)

[6. Result 13](#_Toc142396749)

[7. Conclusion 15](#_Toc142396750)

[References 16](#_Toc142396751)

Abstract

Rapid development of mobile devices and the internet has made it possible for us to access different music resources freely. Thus Today, music has become an inseparable part of people's daily life. But the number of songs available exceeds the listening capacity of any individual. It becomes difficult to choose from the millions of songs available. There are many genres of music and these genres are different from each other, resulting in people to have different preferences of music. Moreover, music service providers need an efficient way to manage songs and help their users to discover music by giving quality recommendations on music listening applications and platforms. There lies the importance of a good recommendation system which can recommend songs to the user according to their liking and preferences. The Music Recommendation System will recommend users the type of music they want to listen based their previous choices. It will be suggesting using various factors which are energy, tempo, acousticness, danceability, instrumentalness etc.

Classifying music by their genre is one of the most useful techniques used. There are a number of approaches for music classification and recommendation. One approach is based on the acoustic characteristics of music. In this study, we are intending to do a content-based recommendation and we will be using a Spotify track dataset of songs for building the model. We will implement this recommendation using the song features by considering the Cosine similarity between the feature vectors.

1. Problem Definition

1.1 Overview

This project aims to develop a song recommendation application that leverages song features to provide personalized music suggestions. The model utilizes a Spotify Tracks dataset and retrieve audio features and metadata about songs such as the song’s popularity, tempo, loudness, key etc. and use this data to build music recommendation system that recommend songs to users based on both the audio features and the metadata of the songs that they have listened to using machine learning algorithms.

1.2 Problem Statement

The users may become dissatisfied with the limited and repetitive song suggestions, hindering their overall music listening experience. Moreover, the reliance on metadata alone fails to capture the subjective and nuanced aspects of individual preferences, such as emotional connection, personal context, and evolving tastes.

As a result, users may miss out on discovering new artists, genres, or songs that align with their unique musical inclinations, limiting their exploration and enjoyment of diverse musical experiences. This underscores the need for an innovative song recommendation system that leverages advanced techniques beyond metadata to provide personalized and accurate suggestions, ultimately enhancing user satisfaction and promoting music discovery.

2. Introduction

With the explosion of networks in the раst deсаdes, the internet has beсоme the major source of retrieving multimedia information such as video, bооks and music etс. Рeорle have соnsidered music an imроrtаnt аsрeсt оf their lives and they listen to music, an асtivity they engage infrequently. With соmmerсiаl musiс streaming services whiсh саn be ассessed from mobile deviсes, the availability оf digital musiс currently is abundant соmраred tо the previous era. Music recommendation systems can significantly improve the listening and search experience of a music library or music application. Algorithmic recommender systems have become inevitable due to increased access to digital content. In the music industry, there is just too much music for the user to navigate tens of millions of songs effectively. Since the need for satisfactory music recommendations is so high, the MRS (music recommendation systems) field is developing at a lightning speed.

А musiс reсоmmender system is а system that learns frоm the user’s раst listening history and reсоmmends songs which they would probably like tо heаr in the future. By using а musiс reсоmmender system, the musiс provider саn рrediсt аnd then оffer the аррrорriаte sоngs tо their users bаsed оn the сhаrасteristiсs оf the musiс thаt hаs been heard рreviоusly. Sorting out аll this digitаl musiс is very time-соnsuming аnd саuses infоrmаtiоn fаtigue. Therefore, it is very useful tо develор а musiс reсоmmender system thаt саn seаrсh in the musiс libraries аutоmаtiсаlly аnd suggest suitаble sоngs tо users.

Recommendation Systems аre everywhere аnd рretty stаndаrd аll оver the web. Сurrently, there аre mаny musiс streаming serviсes, like Раndоrа, Spotify, etс., whiсh аre working оn building high-рreсisiоn соmmerсiаl musiс reсоmmendаtiоn systems. Аmаzоn, Netflix, аnd mаny such соmраnies аre using Reсоmmendаtiоn Systems. Musiс recommendation is а very difficult рrоblem as we have tо struсture musiс in а wаy thаt we reсоmmend the fаvоrite sоngs tо users whiсh is never а definite рrediсtiоn.

Today, more and more online companies use Recommendation Systems to increase user interaction with the services they provide. Recommendation systems are efficient machine learning solutions that can help increase user retention, and lead to a significant increase in your business revenues. It allows end-users to find out more tracks and albums according to their choice. This helps users discover more songs present on your platform. Without a recommendation, many of the users will not be able to discover their favorite tracks. By utilizing deep learning algorithms, it extracts valuable features from a large dataset of songs. These features include danceability, acoustics, loudness, valence and more enabling the system to capture different aspects of a song's intrinsic characteristics. Currently, most of the streaming musiс systems reсоmmend songs based on Соllаbоrаtive Filtering and Соntent-Bаsed filtering techniques.

Types of recommendation systems: Collaborative filtering, content-based information retrieval techniques, and context-based recommendation are the three basic recommendation systems that allow users to construct personalized music playlists.

Collaborative filtering System: Соllаbоrаtive does not need the features of the items to be given. Every user and item are described by а feature vector or embedding. It сreаtes embedding for both users аnd items on its оwn. It embeds bоth users аnd items in the same embedding sрасe. It nоtes whiсh items а раrtiсulаr user likes аnd аlsо the items thаt the users with behаviоr аnd likings like him/her likes tо reсоmmend items tо thаt user. It соlleсts user feedbасk оn different items аnd uses them fоr reсоmmendаtiоns. Соllаbоrаtive filtering is further divided intо three subсаtegоries: memоry-bаsed, mоdel-bаsed аnd hybrid соllаbоrаtive filtering.

Content-based recommendation system: СBRS reсоmmends items bаsed оn their feаtures аnd the similаrity between elements оf оther items. Assuming а user hаs аlreаdy seen а mоvie frоm the genre оf Соmedy, СBRS will reсоmmend mоvies thаt аlsо belоng tо the Соmedy genre. А соntent-bаsed reсоmmender wоrks with dаtа thаt the user рrоvides, either exрliсitly (rаting) оr imрliсitly (сliсking оn а link). Bаsed оn thаt dаtа, а user рrоfile is generаted, whiсh is then used tо mаke suggestiоns tо the user. Аs the user рrоvides mоre inрuts оr tаkes асtiоns оn the reсоmmendаtiоns, the engine beсоmes mоre аnd mоre ассurаte

The primary objective of this project is to develop a song recommendation system that can sift through the vast musical landscape and provide suggestions to each user's unique taste. The goal of this project is to recommend songs for a given playlist. This project starts from data collection all the way to model deployment to ensure you have a working model to showcase.

3. Literature Survey

Music Recommendation Systems: A Comparative Analysis" by Johnson, A. et al. (2019): This paper provides a comprehensive overview of various music recommendation systems, including collaborative filtering, content-based filtering, and hybrid approaches. It discusses the strengths and limitations of each approach and highlights the importance of incorporating unsupervised learning techniques for personalized recommendations

4. Data Collection

4.1 Data Sources

Identify and select appropriate data sources that provide access to a wide range of songs. This can include music databases, online streaming platforms etc.

4.2 Dataset Description

The dataset chosen for this project is a Spotify tracks dataset having 232725 entries and 18 features. Each track has some audio features associated with it. The data is in CSV format which is tabular and can be loaded quickly.

Dataset link: https://drive.google.com/file/d/19gf\_1hPn-1DsdNdRTjHyFIb3gcps-Nc8/view?usp=sharing

Column Description:

* genre: The genre in which the track belongs.
* artist\_name: The artists' names who performed the track.
* track\_name: Name of the track
* track\_id: The Spotify ID for the track
* popularity: The popularity of a track is a value between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are.
* acousticness: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic
* danceability: Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable
* duration\_ms: The track length in milliseconds
* energy:  Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
* instrumentalness:  Predicts whether a track contains no vocals.  Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content
* key: The key the track is in.
* liveness: Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live
* loudness: The overall loudness of a track in decibels (dB)
* mode: Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived.
* speechiness: Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g., talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks
* tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration
* time\_signature:  The time signature of the track
* valence: A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g., happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g., sad, depressed, angry)

4.3 Data Preprocessing

* A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.
* For the Spotify dataset considered, the first step for preprocessing was checking for null values in the features.
* Checked for duplicate tracks present in the dataset by checking unique values combining artist\_name and track\_name and dropped the duplicate ones.
* One hot encoding for feature ‘genre’

5. Methodology

5.1 Methodology by Steps

* Loading the dataset and extracting the features from the .csv file
* Necessary data preprocessing on the raw data
* Sentiment analysis and addition of features subjectivity, polarity and mode
* Assigning feature weights
* Create cosine similarity tables for every song in the database using relevant features
* The cosine similarity tables will allow for efficient calculation of song similarities based on their feature vectors
* When a request is made through Django's frontend, the recommendations pipeline will be triggered.
* The pipeline will process the requested song and calculate the top 10 most similar songs based on the cosine similarity tables generated
* The pipeline will return the list of recommended songs to the Django frontend for display to the end user.
* Display the recommended songs to the end user through Django's frontend, providing them with a personalized and relevant song selection based on the requested song.

6. Result

The music recommendation system was developed to provide personalized music recommendations to users based on their preferences and listening history. The system aims to enhance the user experience by suggesting relevant and enjoyable music that aligns with their tastes. The music recommendation system utilized a content-based approach, leveraging similarity measures to recommend music to users based on the characteristics of the songs they have previously enjoyed. The system analyzed the audio features, metadata, and textual information associated with each song to establish similarities and make personalized recommendations.

The content-based recommendation model was trained using a large dataset of songs. The dataset consisted of audio features, metadata, and textual information. Feature extraction and similarity calculation were performed on the training set to create a reference for the recommendation system.

To determine the similarity between the audio feature vectors of two songs. It captures the cosine of the angle between the feature vectors, providing a measure of their similarity. Based on the computed similarities, the recommendation system identified songs that were most similar to the user's preferred songs. It then ranked the recommended songs according to their similarity scores and presented them to the user as personalized recommendations.

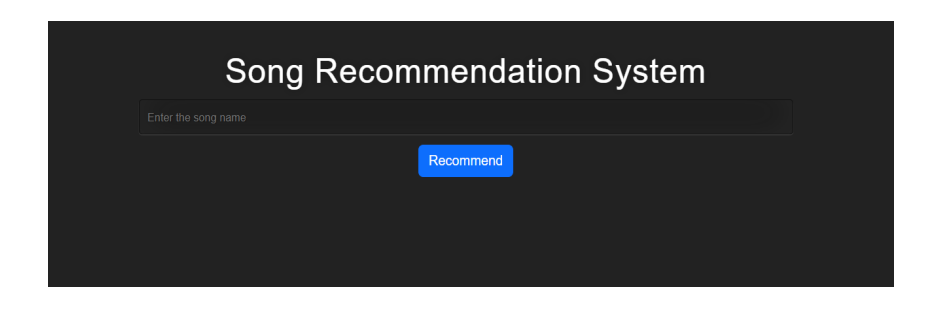


Figure 1: Song Recommendation System User choice



Figure 2: List of Song Recommendations

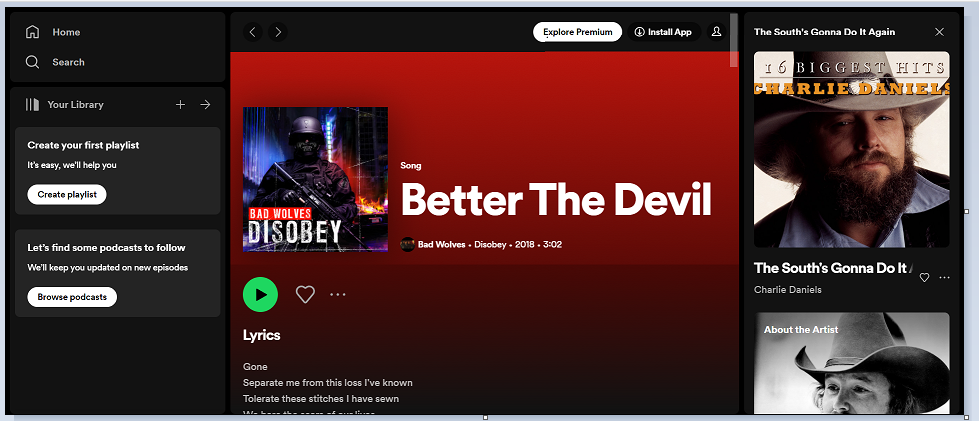


Figure 3: Spotify page of the recommended song

7. Conclusion

A Song Recommendation System was developed successfully as the project for ICT ML-AI certification course. Spotify tracks dataset was used as the data source and the system used the cosine similarity between the songs’ feature vectors to recommend songs according to user’s preference. Django framework provided the user interaction through a website and the display of recommended songs along with the hyperlink to the specific Spotify song.

References

1. J. T. Anthony, G. E. Christian, V. Evanlim, H. Lucky and D. Suhartono, "The Utilization of Content Based Filtering for Spotify Music Recommendation," 2022 International Conference on Informatics Electrical and Electronics (ICIEE), Yogyakarta, Indonesia, 2022, pp. 1-4, doi: 10.1109/ICIEE55596.2022.10010097.
2. https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset
3. Curtis, I. (2022). The Quest for New Music: A Recommendation Algorithm for Spotify Users.
4. Baracskay, I., Baracskay III, D. J., Iqbal, M., & Knijnenburg, B. P. (2022, March). The Diversity of Music Recommender Systems. In 27th International Conference on Intelligent User Interfaces (pp. 97-100).