Music Recommender System Documentation

Introduction

This document outlines the design, implementation, and evaluation of a music recommender system. The system utilizes natural language processing techniques and cosine similarity to recommend songs to users based on their preferences.

Dataset Description

- **Size**: 57,650 rows
- **Columns**:
- Artist: Name of the artist
- Song: Title of the song
- Link: URL link (dropped)
- Text: Text data (processed)

Data Preprocessing

- **Dropped Columns**: The 'link' column was dropped as it was deemed unnecessary for the recommendation process.
- **Null Values and Duplicates**: No null values or duplicates were found in the dataset.
- **Sampling**: A sample of 15,000 rows was selected from the original dataset to reduce computational complexity.
- **Regular Expressions**: Removed from the text data to clean and preprocess it.
- **Tokenization and Stemming**: Utilized PorterStemmer and word_tokenize to tokenize and stem the text data.

Feature Engineering

- **TF-IDF Vectorization**: Utilized TfidfVectorizer to convert the text data into numerical vectors, accounting for term frequency and inverse document frequency.

Recommender System

- **Cosine Similarity**: Employed cosine similarity to compute the similarity between songs based on their TF-IDF vectors.
- **Recommendation Generation**: For a given song, the system recommends similar songs based on their cosine similarity scores.

Evaluation

- **Methodology**: Evaluated the recommender system using metrics such as precision, recall, and F1-score.
- **Validation**: Validated the recommendations by comparing them with user feedback and preferences.
- **Performance**: Assessed the performance of the system in terms of recommendation quality and computational efficiency.

Conclusion

The music recommender system demonstrates the effectiveness of natural language processing techniques and cosine similarity in recommending songs to users based on their preferences. Further improvements and optimizations can be explored to enhance the accuracy and scalability of the system.