

UNIVERSIDADE ANHEMBI MORUMBI

Yuri Freire Kuramoto
Aron Uliani dos Santos

Projeto de Sistema de Recomendação de Músicas.

São Paulo

2024

RESUMO

Este relatório apresenta o desenvolvimento de um sistema de recomendação de músicas, utilizando técnicas de machine learning e análise de dados. Por meio do processamento de dados, vetorização de texto, normalização de atributos numéricos e cálculo de similaridade, criamos um modelo eficiente para recomendar músicas baseando-se em características textuais e numéricas.

Palavras-chave: Música, Similaridade, Machine Learning.

SUMÁRIO

1.Introdução

2.Base de Dados

2.1. Motivação e Escolha da Base

2.2. Estrutura da Base de Dados

3.Análise e Modelagem de Dados

3.1. Pré-processamento e Limpeza

3.2. Desenvolvimento do Modelo

3.3. Avaliação do Modelo

4.Conclusão

1. INTRODUÇÃO

O sistema de recomendação de músicas desenvolvido neste projeto busca auxiliar usuários a descobrir músicas novas baseadas em suas preferências. A proposta inclui a criação de um modelo que combina atributos textuais (como gênero e artistas) e numéricos (como dançabilidade, energia e popularidade) para calcular a similaridade entre músicas e sugerir recomendações relevantes.

2. BASE DE DADOS

2.1. Motivação e Escolha da Base

A base de dados utilizada foi extraída de um repositório público e contém informações detalhadas sobre músicas, como título, artistas, gênero, e características acústicas. A escolha dessa base foi motivada pela sua riqueza em informações e pelo potencial de aplicação prática em sistemas de streaming e playlists personalizadas.

2.2. Estrutura da Base de Dados

As principais colunas utilizadas no projeto incluem:

- Título da Música (track_name): Nome da música.
- Artistas (artists): Nome(s) do(s) artista(s) responsável(eis).
- Gênero (track_genre): Gênero associado à música.
- Atributos Numéricos: danceability, energy, valence, tempo, popularity.

Esses dados foram processados e combinados para criar uma matriz de características representativa de cada música.

3. ANÁLISE E MODELAGEM DE DADOS

3.1. Pré-processamento e Limpeza

O pré-processamento dos dados incluiu:

- Tratamento de valores nulos: Substituição de valores ausentes por categorias padrão ou medianas.
- Combinação de atributos textuais em uma única coluna, para melhorar o desempenho do modelo de similaridade.
- Normalização dos atributos numéricos entre 0 e 1.

3.2. Desenvolvimento do Modelo

O modelo foi implementado em Python, utilizando as seguintes etapas:

1. Vetorização de Atributos Textuais:

- Foi utilizado o TF-IDF (Term Frequency - Inverse Document Frequency) para transformar os textos em vetores numéricos.

2. Normalização de Atributos Numéricos:

- Foi aplicado o escalonamento com MinMaxScaler.

3. Combinação e Redução Dimensional:

- Os dados textuais e numéricos foram combinados em uma matriz única, que foi reduzida para 100 dimensões usando PCA (Principal Component Analysis).

4. Cálculo de Similaridade:

- A similaridade de cosseno foi usada para medir o quão próximas as músicas estão entre si.

3.3. Avaliação do Modelo

Foram implementadas métricas de qualidade para avaliar as recomendações:

1. Similaridade Média: Mede a proximidade das músicas recomendadas com a música de entrada.
2. Diversidade de Gêneros: Avalia a variedade de gêneros presentes nas recomendações.
3. Métricas obtidas para uma música de exemplo:
 - Similaridade Média: 0,85
 - Diversidade de Gêneros: 3 gêneros distintos

4. CONCLUSÃO

O projeto demonstrou como um sistema de recomendação pode ser construído de forma eficiente usando dados textuais e numéricos. A integração de técnicas de processamento de texto, escalonamento numérico e cálculo de similaridade resultou em um modelo capaz de gerar recomendações relevantes e diversas.