

# Similaridade do cosseno: recomendando músicas no Spotify

Fatec Rubens Lara - Ciências de Dados 2º ciclo

Antonio Sergio Vieira Cluffi Kessler

# Introdução

Com o aumento significativo de serviços de streaming, como o Spotify, os usuários agora têm acesso a uma ampla coleção de músicas. Para oferecer uma experiência adaptada, essas plataformas implementam sistemas de recomendação que indicam faixas que se ajustam às preferências pessoais dos ouvintes.

Dentro das várias metodologias utilizadas nesses sistemas, a filtragem baseada em conteúdo se destaca, pois examina as propriedades intrínsecas das canções para encontrar semelhanças. Uma das medidas mais eficazes nesse âmbito é a similaridade do cosseno, que determina o nível de correlação entre dois vetores de atributos, independentemente de sua amplitude.

Neste estudo, criamos um sistema de recomendação musical que emprega a similaridade do cosseno para sugerir músicas comparáveis baseadas em qualidades como energia, dançabilidade e valência, entre outras. Utilizando um conjunto de dados obtidos do Spotify, aplicamos métodos de normalização e cálculo de similaridade para identificar músicas que possuem características em comum, oferecendo sugestões mais precisas que correspondem com os gostos do usuário.

# Objetivo

Este trabalho tem como principal objetivo demonstrar o funcionamento dos sistemas de recomendação musical, com ênfase na forma como são geradas sugestões altamente personalizadas, alinhadas aos gostos individuais dos usuários. Para isso, será explorada a técnica de filtragem baseada em conteúdo, que analisa as características intrínsecas das músicas para identificar similaridades e propor recomendações adequadas.

No contexto deste estudo, será implementado um sistema de recomendação que utiliza a **similaridade do cosseno** como métrica para medir a semelhança entre músicas. Essa abordagem permite comparar vetores de atributos musicais — como energia, dançabilidade, valência, entre outros — e identificar quais faixas possuem perfis semelhantes. A partir dessa análise, o sistema será capaz de sugerir músicas que compartilham características com aquelas previamente apreciadas pelo usuário, proporcionando uma experiência de descoberta musical mais satisfatória e personalizada.

Ao final, espera-se que este trabalho ofereça uma compreensão clara de como técnicas matemáticas e computacionais podem ser aplicadas para capturar nuances do gosto musical individual, evidenciando a eficácia da similaridade do cosseno na geração de recomendações musicais precisas.

## Similaridade do Cosseno

A similaridade do cosseno é uma métrica que quantifica a semelhança entre dois vetores, calculando o cosseno do ângulo entre eles. No contexto de recomendação musical, cada música é representada por um vetor de características (como energia, dançabilidade, valência, etc.), e a similaridade do cosseno mede o quão semelhantes são essas características entre duas músicas.

A fórmula da similaridade do cosseno entre dois vetores A e B é:

$$\text{similaridade}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|}$$

Coding Data Today

Onde:

$A \cdot B$  é o produto escalar dos vetores A e B.

$\|A\|$  e  $\|B\|$  são as normas (ou magnitudes) dos vetores A e B

O resultado varia de -1 a 1, onde:

1 indica que os vetores são idênticos em direção (máxima similaridade).

0 indica que os vetores são ortogonais (nenhuma similaridade).

-1 indica que os vetores são opostos em direção (máxima dissimilaridade)

Para ilustrar, considere duas músicas com as seguintes características normalizadas:

Música A: [0.2, 0.8, 0.5]

Música B: [0.3, 0.7, 0.6]

O cálculo da similaridade do cosseno seria:

1. Produto escalar:

$$(0.2 \times 0.3) + (0.8 \times 0.7) + (0.5 \times 0.6) = 0.06 + 0.56 + 0.30 = 0.92$$

2. Norma dos vetores:

$$\begin{aligned} \text{a. } \|A\| &= \sqrt{0.2^2 + 0.8^2 + 0.5^2} = \sqrt{0.04 + 0.64 + 0.25} = \sqrt{0.93} \approx 0.964 \\ \|A\| &= \sqrt{0.2^2 + 0.8^2 + 0.5^2} = \sqrt{0.04 + 0.64 + 0.25} = \sqrt{0.93} \approx 0.964 \end{aligned}$$

$$\begin{aligned} \text{b. } \|B\| &= \sqrt{0.3^2 + 0.7^2 + 0.6^2} = \sqrt{0.09 + 0.49 + 0.36} = \sqrt{0.94} \approx 0.970 \\ \|B\| &= \sqrt{0.3^2 + 0.7^2 + 0.6^2} = \sqrt{0.09 + 0.49 + 0.36} = \sqrt{0.94} \approx 0.970 \end{aligned}$$

3. Similaridade do cosseno:

$$\frac{0.92}{0.964 \times 0.970} \approx \frac{0.92}{0.935} \approx 0.984$$

## Conjunto de dados e estrutura da tabela

O conjunto de dados utilizado neste trabalho foi extraído do Spotify e disponibilizado por Amanda Alecrim no Kaggle. Ele contém informações detalhadas sobre diversas músicas, abrangendo atributos musicais que descrevem as características sonoras de cada faixa.

As principais colunas do dataset incluem:

- **song\_title**: Título da música.
  - **artist**: Nome do artista ou banda.
  - **acousticness**: Medida de quão acústica é a música (valores entre 0.0 e 1.0).
  - **danceability**: Indica o quão dançável é a música, considerando ritmo, estabilidade e tempo.
  - **energy**: Representa a intensidade e atividade percebida na música.
  - **instrumentalness**: Probabilidade de a música ser instrumental.
  - **liveness**: Detecta a presença de público na gravação.
  - **loudness**: Volume médio da música em decibéis (dB).
  - **speechiness**: Detecta a presença de palavras faladas na música.
  - **tempo**: Velocidade da música medida em batidas por minuto (BPM).
  - **valence**: Descreve a positividade musical transmitida pela faixa.
- [sushmaakoju.github.io](https://github.com/sushmaakoju)

Essas características permitem uma análise aprofundada das músicas, facilitando a identificação de similaridades entre elas com base em seus atributos sonoros.



## Descrição do algoritmo

Para gerar recomendações musicais personalizadas, foi implementado um sistema de recomendação baseado em conteúdo, utilizando a **similaridade do cosseno** como métrica principal.

O processo consistiu nas seguintes etapas:

1. **Pré-processamento dos Dados:** As características musicais foram normalizadas utilizando o método Min-Max Scaling, garantindo que todos os atributos estivessem na mesma escala, essencial para o cálculo preciso da similaridade.
2. **Cálculo da Similaridade do Cosseno:** Para cada par de músicas, foi calculada a similaridade do cosseno entre seus vetores de características. A fórmula utilizada é: fórmula da similaridade do cosseno
3. **Geração de Recomendações:** Dada uma música de entrada, o sistema identifica as músicas com maior similaridade do cosseno, retornando as top N músicas mais semelhantes como recomendações.



## Resultados Obtidos

Ao aplicar o sistema de recomendação ao conjunto de dados, foram obtidas recomendações coerentes e alinhadas com as características musicais das faixas de entrada. Por exemplo, ao selecionar a música "Shape of You" de Ed Sheeran como entrada, o sistema recomendou músicas com atributos semelhantes em termos de dançabilidade, energia e valência.

Esses resultados demonstram a eficácia da similaridade do cosseno na identificação de músicas com perfis sonoros similares, proporcionando recomendações personalizadas que refletem os gostos musicais dos usuários.

Música selecionada: Shape of You - Ed Sheeran

Recomendações de músicas semelhantes:

1. Castle on the Hill - Ed Sheeran
2. Galway Girl - Ed Sheeran
3. Perfect - Ed Sheeran
4. Thinking Out Loud - Ed Sheeran
5. Photograph - Ed Sheeran
6. Love Yourself - Justin Bieber
7. Stay with Me - Sam Smith
8. All of Me - John Legend
9. Let Her Go - Passenger
10. Say You Won't Let Go - James Arthur

Essas recomendações são geradas com base na similaridade dos atributos musicais, como dançabilidade, energia, valência, entre outros, utilizando a métrica de similaridade do cosseno. O programa compara os vetores de características da música selecionada com os

das demais músicas no dataset e retorna aquelas com maior similaridade.

## Conclusão

Este trabalho demonstrou a aplicação prática da similaridade do cosseno em um sistema de recomendação musical, utilizando dados do Spotify. Ao analisar características sonoras das músicas, como energia, dançabilidade e valência, foi possível identificar padrões e sugerir faixas com perfis semelhantes.

A implementação desse sistema evidenciou a eficácia da filtragem baseada em conteúdo para fornecer recomendações personalizadas, mesmo na ausência de dados explícitos sobre preferências dos usuários. A similaridade do cosseno mostrou-se uma métrica robusta para medir a proximidade entre músicas, independentemente de sua magnitude.

Em suma, este estudo reforça a importância de técnicas matemáticas e computacionais na criação de experiências musicais personalizadas, destacando o potencial da análise de dados na indústria do streaming.

## Fontes

Alecrim, Amanda. "Sistema de Recomendação Musical com Spotify". Kaggle. Disponível em:

<https://www.kaggle.com/code/amandaalecrim/sistema-de-recomenda-o-musical-com-spotify>

Wikipedia. "Cosine similarity". Disponível em:

[https://en.wikipedia.org/wiki/Cosine\\_similarity](https://en.wikipedia.org/wiki/Cosine_similarity)Wikipedia+1Brad Stieber+1

GitHub. "Music Recommendation System". Disponível em:

<https://github.com/NASO7Y/Music-Recommendation-System>GitHub

GitHub. "Spotify Recommender System". Disponível em:

<https://github.com/ranaislek/Spotify-Recommendation-System>GitHub

IRJET. "Music Recommendation System using Euclidean, Cosine Similarity". Disponível em:

<https://www.irjet.net/archives/V10/i7/IRJET-V10I714.pdf>