

# **Relatório Da Criação Do Código De Recomendação**

## **Utilizando A Similaridade Do Cosseno**

### **Objetivo**

O objetivo deste trabalho foi desenvolver um algoritmo de recomendação baseado no conceito de similaridade do cosseno, abordado nas aulas de Álgebra Linear. Esse conceito é amplamente utilizado no processamento de texto para comparar documentos ou frases representados por vetores, como vetores de frequência de palavras, por exemplo.

A similaridade do cosseno permite medir o grau de semelhança entre dois textos, considerando apenas a direção dos vetores e desconsiderando seu tamanho (magnitude). A fórmula para calcular a similaridade do cosseno entre dois vetores A e B é:

$$\text{Similaridade}(A, B) = (A \cdot B) / (||A|| \times ||B||)$$

onde:

- $A \cdot B$  é o produto escalar entre os vetores;
- $||A||$  e  $||B||$  são as normas (magnitudes) dos vetores

Se a similaridade do cosseno for 1, os textos são iguais ou muito semelhantes, próxima de 0, os textos são bastante diferentes. Quanto maior o valor (mais próximo de 1), maior a semelhança. Quanto menor (mais próximo de 0), menor a semelhança. Essa métrica permite comparar dois textos de forma matemática e objetiva, mesmo que eles não sejam idênticos palavra por palavra.

## Tratamento dos dados

O *dataset* utilizado neste trabalho, intitulado Letras de Bossa Nova e retirado da plataforma *Kaggle*, reúne letras de músicas do gênero Bossa Nova, juntamente com informações como nome da música, artista, letra da música, compositores e idioma da música (português, inglês, espanhol, italiano, entre outros).

Como o foco da análise está na língua portuguesa, foi feito um tratamento inicial para remover todas as músicas que não estavam nesse idioma. O *dataset* original continha 6.107 músicas, mas, após a filtragem, passou a contar com 4.118 músicas em português. Esse processo foi essencial para garantir que os resultados do algoritmo de recomendação, baseado na similaridade do cosseno, fossem mais consistentes e relevantes.

## Algoritmos Desenvolvidos

Com o intuito de investigar diferentes metodologias para sistemas de recomendação, foram implementados dois algoritmos distintos que utilizam a similaridade do cosseno como métrica de comparação:

- Recomendação fundamentada no título da música: nessa abordagem, os títulos das canções foram transformados em vetores e empregados como base para a mensuração da similaridade entre as músicas. Essa estratégia visa identificar faixas com títulos semelhantes, o que pode refletir proximidade temática ou estilística.
- Recomendação fundamentada na letra da música: neste segundo método, utilizam-se as letras completas das músicas para o cálculo da similaridade. Tal procedimento tende a gerar resultados mais consistentes, pois incorpora o conteúdo lírico integral, permitindo recomendações mais precisas em relação a temas, sentimentos e vocabulário presentes nas composições.

Ambos os algoritmos aplicam técnicas de vetorização textual para converter os dados textuais em representações numéricas, possibilitando a aplicação do cálculo da similaridade do cosseno entre os diferentes elementos.

### **Resultados:**

Para validar os algoritmos implementados, foram realizados testes práticos com a inserção de trechos reais, permitindo observar o comportamento de cada modelo. As duas imagens a seguir ilustram os resultados produzidos por ambas as abordagens de recomendação: uma baseada no nome da música e outra fundamentada no conteúdo lírico.

Na primeira imagem, visualizamos os retornos obtidos pelo modelo que opera a partir do título das músicas, recomendando os 10 títulos mais próximos aos inseridos. Confira a seguir:

```
Digite o nome da música para recomendar: girassol
O Girassol - Vinicius de Moraes
Catavento e Girassol - Sérgio Mendes
Tomara - Alaide Costa
Tempo Calado - Alaide Costa
Quem Sou Eu? - Alaide Costa
Diariamente - Alaide Costa
Coração - Alaide Costa
Absinto - Alaide Costa
Amigo Amado - Alaide Costa
Onde Está Você - Alaide Costa
```

Já na segunda imagem, observa-se o funcionamento do algoritmo que analisa as letras completas das músicas. Por considerar o texto integral, essa abordagem consegue entregar recomendações mais ricas e coerentes com o conteúdo emocional, temático ou linguístico da entrada fornecida.

Digite a letra ou parte da letra da música para recomendar: cheia de graça

Músicas recomendadas:

Título: Mulher Passarinho

Artista: Agostinho dos Santos

Letra (trecho):

Ave maria cheia de graça Foi assim que o menino aprendeu a rezar Como não entendia o que ave maria O que cheia de graça que riam ...

Título: Salve Regina

Artista: Benedictus

Letra (trecho):

Maria, doce mãe Mostra-nos o caminho O caminho do teu filho Jesus Um anjo declarou: Ave Maria Que tu seria cheia de graça E do t...

Título: Januária

Artista: Vinicius de Moraes

Letra (trecho):

Toda gente homenageia Januária na janela Até o mar faz maré cheia Pra chegar mais perto dela O pessoal desce na areia E bat uca p...

Título: Januária

Artista: Dori Caymmi

Letra (trecho):

Toda gente homenageia Januária na janela Até o mar faz maré cheia Pra chegar mais perto dela O pessoal desce na areia E bat uca p...

Título: Canção Bonita

Artista: Simone Pelissari

Letra (trecho):

Quem cantou aquela canção bonita Que me fez lembrar você Por favor me dê uma pista Impossível descrever porque a canção me fez s...

Ativar o Wii

Título: Garota de Ipanema

Artista: Os Cariocas

Letra (trecho):

Olha que coisa mais linda, Mais cheia de graça. É ela a menina que vem e que passa, num doce balanço a caminho do mar. Moça do c...

Título: Garota de Ipanema

Artista: Raphael Rabello

Letra (trecho):

Olha que coisa mais linda Mais cheia de graça É ela menina Que vem e que passa Num doce balanço A caminho do mar Moça do co rpo d...

Título: Garota de Ipanema

Artista: Vinicius de Moraes

Letra (trecho):

Olha que coisa mais linda, mais cheia de graça É ela, menina, que vem e que passa Num doce balanço, caminho do mar Moça do corpo...

Título: Garota de Ipanema

Artista: Roberto Menescal

Letra (trecho):

Olha que coisa mais linda Mais cheia de graça É ela menina que vem e que passa Num doce balanço a caminho do mar Moça do co rpo d...

Título: Garota de Ipanema

Artista: Baden Powell

Letra (trecho):

Olha que coisa mais linda Mais cheia de graça É ela menina que vem e que passa Num doce balanço a caminho do mar Moça do co rpo d...

## **Conclusão**

Com a realização deste projeto, foi possível aplicar, de maneira prática, alguns dos conceitos fundamentais da Ciência de Dados, como o pré-processamento de dados, a vetorização de textos e a medição de similaridade entre informações. Utilizando a técnica de similaridade do cosseno, foram desenvolvidos dois algoritmos de recomendação: um baseado no título da música e outro no conteúdo da letra. Essa experiência mostrou como diferentes formas de representação textual impactam os resultados, reforçando a importância da escolha adequada das variáveis na análise de dados.

Além disso, o projeto ajudou a compreender o fluxo de um processo típico em Ciência de Dados, desde a coleta e tratamento do dataset até a análise dos resultados. Mesmo sendo um trabalho inicial e com limitações, os resultados obtidos reforçaram o potencial do uso de técnicas de processamento de linguagem natural (PLN) para gerar valor a partir de dados textuais. Assim, esta atividade contribuiu para o desenvolvimento das habilidades analíticas e técnicas essenciais para quem está começando na área de Ciência de Dados.