

Vector Databases

Sistemas de Recomendação Inteligentes

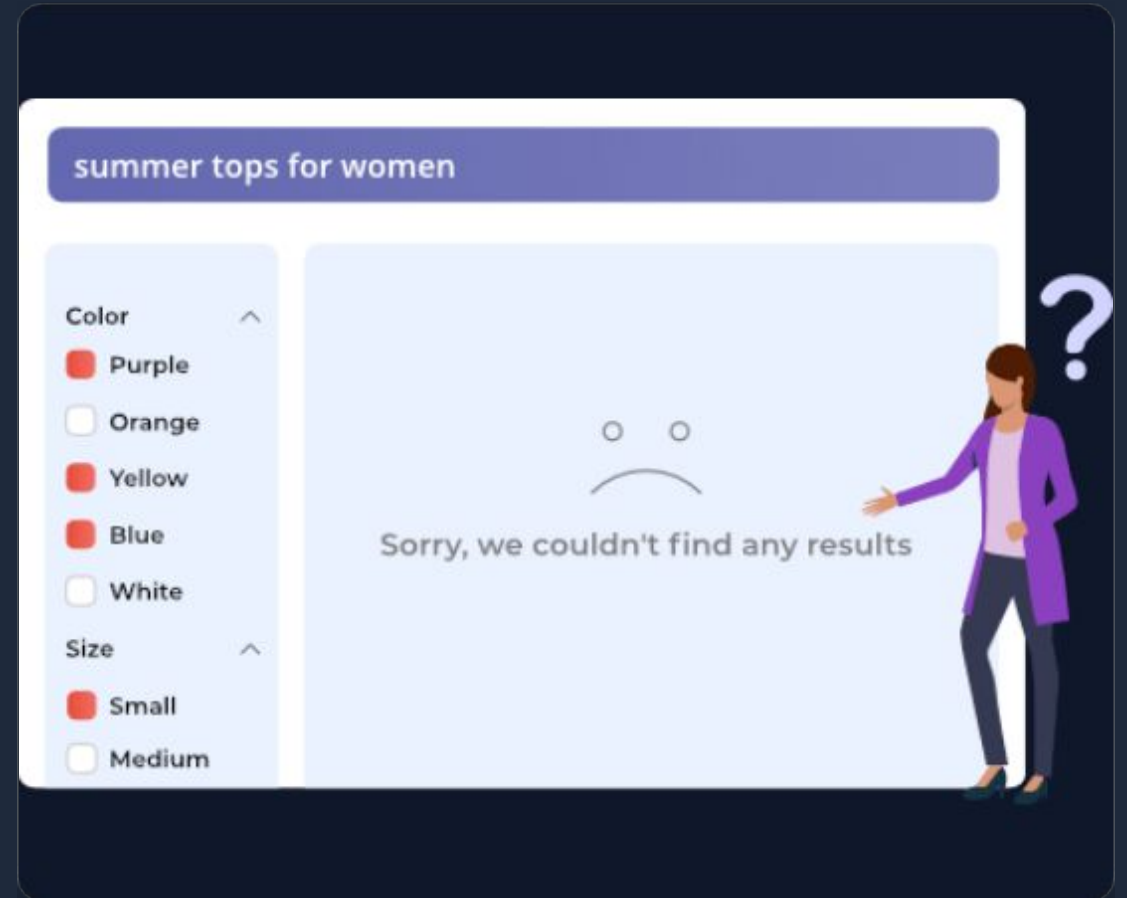
INTEGRANTES DO GRUPO

- Diego Lemos Passani Gomes
- Lucas Malentachi de Sousa
- Nicolly de Paula Santos
- Miguel Fernandes Bellini

O Desafio da Busca Literal

A Falha da Palavra-Chave

- ✗ **Limitação Semântica:** Bancos de dados tradicionais (SQL) buscam por correspondência exata de caracteres.
- ✗ **O Erro Prático:** Ao buscar *"filme de medo"*, o sistema ignora descrições contendo *"terror"* ou *"horror"* se a palavra exata não existir.
- ✗ **Experiência do Usuário:** Isso resulta em "Zero Resultados" ou recomendações irrelevantes (como Toy Story), frustrando o usuário final.



A Solução Vetorial

De Palavras para Matemática

Substituímos a busca textual pela **Busca Semântica** usando Embeddings.

Como funciona:

- Textos são convertidos em coordenadas numéricas (vetores).
- Conceitos similares ("Medo" e "Aterrorizante") ficam próximos matematicamente.
- A busca mede a *distância* entre a intenção do usuário e o conteúdo.



Implementação com ChromaDB

Utilizamos Python e ChromaDB para criar um pipeline de recomendação inteligente que entende o contexto, não apenas a sintaxe.



1. Vetorização

As sinopses dos filmes são processadas pelo modelo *all-MiniLM*, transformando textos descritivos em listas de números (embeddings) armazenadas na memória.



2. Cálculo de Distância

A query do usuário ("filme que me assuste") é vetorizada. O sistema calcula a proximidade cosseno contra todos os filmes do banco.



3. Recomendação

O filme com a menor distância (Ex: 0.52 para "It") é retornado. Resultados distantes (Ex: 1.51 para "Titanic") são descartados.