

FACULDADE DE TECNOLOGIA BAIXADA
SANTISTA

CIÊNCIA DE DADOS

Ana Luiza Fonseca Gonçalves

Recomendação de Livros: Similaridades dos
Cossenos

SANTOS – SP

2025

1 Introdução

Este trabalho tem como foco a recomendação de livros utilizando a técnica de similaridade de cossenos na Skoob. A similaridade de cossenos é uma métrica que mede a semelhança entre dois vetores, permitindo identificar quais livros são mais semelhantes entre si com base em suas características. Essa abordagem é especialmente útil em sistemas de recomendação, onde o objetivo é sugerir itens que possam interessar ao usuário com base em suas preferências anteriores e benefício para a plataforma como a validação do acervo disponível.

2 Descrição do Dataset

O dataset utilizado neste trabalho é o *Livros Skoob*, que foi retirado do Kaggle e renomeado `livros.csv`. Este conjunto de dados foi extraído da plataforma de leitura e avaliação de livros, o Skoob, e contém informações relevantes sobre os livros mais lidos na plataforma pelos usuários. As principais características do dataset incluem:

- **Título do Livro:** Nome do livro.
- **Autor:** Nome do autor do livro.
- **Gênero:** Categoria ou gênero literário ao qual o livro pertence.
- **Avaliações:** Notas dadas pelos usuários, que refletem a popularidade e a recepção do livro.
- **Descrição:** Breve descrição do conteúdo do livro.

Essas informações são fundamentais para calcular a similaridade entre os livros e, assim, gerar recomendações personalizadas e reavaliação do acervo por parte da plataforma.

3 Objetivo

O principal objetivo deste trabalho é desenvolver um sistema de recomendação de livros que utilize a similaridade de cossenos para sugerir livros a partir das preferências dos usuários. Através da análise dos títulos dos livros é buscada a descrição deles, e a partir dela é feito o cálculo da similaridade dos cossenos, com o intuito de identificar quais livros são mais semelhantes e, portanto, mais propensos a serem do interesse de um usuário específico ou ainda, avaliar a atualização da plataforma quanto ao acervo disponível.

4 Resultados Obtidos

A aplicação da similaridade do cosseno sobre vetores TF-IDF dos campos textuais (como título, autor e descrição) resultou em um sistema de recomendação eficaz, capaz de identificar livros com alta similaridade temática. A seguir, apresenta-se uma análise mais detalhada dos resultados, incluindo aspectos técnicos da implementação e visualizações.

Implementação Técnica

A implementação do modelo foi realizada utilizando as seguintes etapas:

- **Carregamento do Dataset:**

O dataset `livros.csv` é carregado utilizando a biblioteca `pandas`.

```
df = pd.read_csv('livros.csv')
```

- **Pré-processamento dos Dados:**

Padronização dos títulos para comparação e remoção de duplicatas.

```
df['titulo_lower'] = df['titulo'].str.strip().str.lower()
df = df.drop_duplicates(subset='titulo_lower', keep='first').reset_index(drop=True)
```

- **Tratamento de Texto:**

Remoção de pontuações e stopwords. Além da modificação do dataset, nele foi adicionado mais uma coluna que armazenará os textos *tratados*.

```
def tratamento_texto(texto):
    texto = str(texto).lower()
    texto = texto.translate(str.maketrans('', '', string.punctuation))
    tokens = word_tokenize(texto)
    stop_words = set(stopwords.words("portuguese"))
    tokens = [palavra for palavra in tokens if palavra not in stop_words]
    return ' '.join(tokens)
```

```
df["descricao_trat"] = df["descricao"].apply(tratamento_texto)
```

- **Vetorização das Descrições:**

Utilização da biblioteca `TfidfVectorizer` para transformar as descrições dos livros em vetores TF-IDF.

```
vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform(descriptions)
```

- **Cálculo da Similaridade do Cosseno:**

Cálculo da similaridade entre os vetores TF-IDF.

```
sim_scores = cosine_similarity(tfidf_matrix[idx], tfidf_matrix).flatten()
```

```
similar_books = pd.DataFrame({
    'Livro': titles,
    'Editora': editoras,
    'Autor(a)': autor,
    'Similaridade': sim_scores
})
```

- **Geração de Recomendações:**

Montagem de um `DataFrame` com os livros recomendados, suas editoras, autores e a similaridade calculada.

O resultado aparece através de uma interface intuitiva. Nela o usuário digita o título do livro que deseja buscar a similaridade (Figuras 1, 2 e 3).



Figura 1: Interface do usuário para digitação do título do livroFonte: Autor (2025)

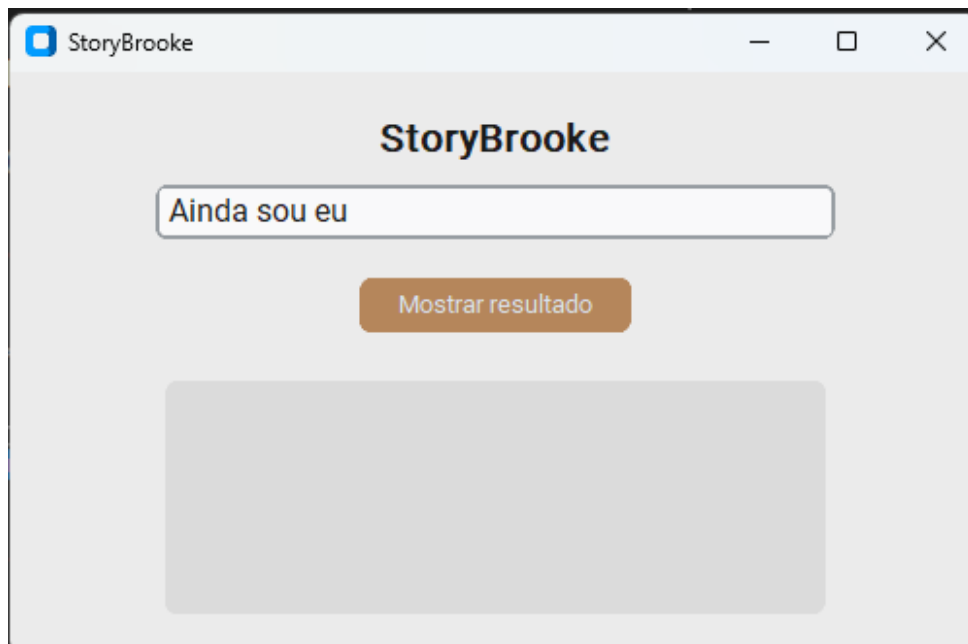


Figura 2: Foi adicionado o nome do livroFonte: Autor (2025)

Livro	Editora	Autor(a)	Similaridade
Como eu era antes de você	Intrínseca	Jojo Moyes	0.3605330014268985
Depois de você	Intrínseca	Jojo Moyes	0.2553819216307244
O Poço dos Desejos	Rocco	David Baldacci	0.19862072764739855
O Substituto	Intrínseca	David Nicholls	0.1579628945405086
Humana, Demasiado Humana	Rocco	Luziá	0.14966878515433002
O Assassino em Mim	Planeta	Jim Thompson	0.14120896249542275
Inocência, conhecimento e encantamento	Editora	Osho	0.12670243487766664
Cama	Rocco	David Whitehouse	0.12040678781548772
Ela Foi Até o Fim	Galera	Meg Cabot	0.11812011577264935
Abandono	Record	Meg Cabot	0.11688371405155443
Por trás de seus olhos	Intrínseca	Sarah Pinborough	0.11465003344238436
O ritmo da vida	Sextante	Matthew Kelly	0.10938519233786204

Figura 3: Resultado finalFonte: Autor (2025)

5 Conclusão

Podemos concluir que a análise da similaridade dos cossenos aplicada ao dataset da plataforma Skoob permitiu identificar os livros mais semelhantes ao título inserido pelo usuário. Conforme os resultados obtidos, o primeiro livro apresentado na tabela possui a maior similaridade, enquanto o último apresenta a menor. A limitação de exibição para apenas 12 resultados se deve ao grande volume de dados disponíveis, garantindo

que o usuário tenha acesso às recomendações mais relevantes. Esse método se mostra eficiente para auxiliar leitores na descoberta de obras que compartilham características semelhantes, ampliando suas opções de leitura.