



Universidade Presbiteriana Mackenzie

Título: Livros em Rede: Desvendando Recomendações Inteligentes

CURSO: Tecnologia em Ciência de dados

POLO DE APOIO PRESENCIAL: Polo EAD SP - Polo EAD Goiânia

SEMESTRE: 04

COMPONENTE CURRICULAR / TEMA: Projeto aplicado III

NOME COMPLETO DOS ALUNOS:

Nome : Gustavo Silva Rios

RA : 10415824

Nome : Silas de Souza Ferreira

RA: 10414793

Nome : Israel Soares do N. Viana

RA: 10414894

Nome : Danilo Brito da Silva

RA: 10415882

Sumário

1	Introdução.....
2	Base de dados utilizada e repositório.....
3
4
5



Introdução

Com a popularidade crescente de plataformas como YouTube, Amazon e Netflix, os sistemas de recomendação se tornaram uma parte diferencial do nosso dia a dia. Esses sistemas, movidos por algoritmos sofisticados, ajudam a direcionar nossas escolhas, seja na hora de assistir a um filme, comprar um produto ou escolher o próximo livro. Em uma época onde a leitura precisa competir com tantas outras formas de entretenimento, a capacidade de recomendar livros que realmente atraiam os leitores pode ser fundamental para incentivar o hábito da leitura.

Nosso estudo explora o dataset Book-Crossing, coletado em 2004, que oferece um panorama das interações entre leitores e livros em uma plataforma que se autodenomina uma "Biblioteca Mundial". O BookCrossing não é apenas um espaço de troca de livros; é uma rede global onde cada livro tem sua história, tanto no conteúdo quanto em sua jornada entre leitores. Analisando esse dataset, buscamos desenvolver um sistema de recomendação capaz de não só sugerir títulos, mas de conectar leitores a livros que realmente fazem a diferença.

No cenário atual, onde o tempo é curto e as opções são muitas, encontrar o livro certo pode transformar a experiência de leitura. Este estudo se baseia na ideia de que a leitura tem um impacto profundo e que, com o apoio da ciência de dados, podemos ajudar a redescobrir esse prazer, especialmente ao sugerir livros alinhados aos gostos e interesses de cada leitor.

Os objetivos deste trabalho são:

- Desenvolver e testar diferentes algoritmos de recomendação aplicados ao dataset Book-Crossing, com foco em personalização.
- Explorar como fatores demográficos, como idade e localização, influenciam as preferências de leitura.
- Avaliar a qualidade dos dados e seu efeito na precisão das recomendações.
- Criar um sistema híbrido de recomendação que combine métodos baseados em conteúdo e filtragem colaborativa, garantindo sugestões mais relevantes.

Com este estudo, buscamos valorizar a leitura, oferecendo recomendações que realmente possam enriquecer a experiência dos leitores. Em um mundo com uma infinidade de opções e tempo cada vez mais escasso, encontrar o livro certo é um desafio. Nosso objetivo é desenvolver um sistema de recomendação que vá além da popularidade dos títulos, conectando leitores a livros que realmente correspondam aos seus gostos e interesses. Ao analisar o comportamento de leitura e fatores como idade e localização, podemos criar sugestões mais personalizadas e relevantes, ajudando a evitar leituras que não atendam às expectativas e promovendo descobertas literárias significativas.



“Um livro não é apenas um amigo, ele faz amizades para você. Quando você possui um livro com mente e alma, você é enriquecido. Mas quando você o passa adiante, você é triplamente enriquecido.”

– Henry Miller
Os Livros da Minha Vida (1969)

Base de dados utilizada e repositório

Os dados utilizados foram extraídos da plataforma Kaggle, que foi coletado por Cai-Nicolas Ziegler durante um período de 4 semanas (agosto/setembro de 2004) na comunidade Book-Crossing, com a permissão de Ron Hornbaker, CTO da Humankind Systems. A base de dados contém informações de 278.858 usuários (anônimos, mas com dados demográficos) que forneceram 1.149.780 avaliações (explícitas/implícitas) sobre 271.379 livros.

```
[76]: display(df_user_ratings.sample(5))
```

	User-ID	ISBN	Book-Rating	Location	Age
1018540	244171	0802130208	0	moline, illinois, usa	51.0
784571	189835	0517092921	5	honolulu, hawaii, usa	NaN
171415	37227	002862825X	6	hamilton, ohio, usa	27.0
740521	178997	3442410665	0	würzburg, bayern, germany	33.0
877409	212740	0553576801	4	fort myers, florida, usa	NaN

```
[78]: display(df_books.sample(5))
```

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher
268803	9053330313	De dikke man	Ischa Meijer	1991	Prometheus
162977	0835605612	Shaman and the Medicine Wheel	Evelyn Eaton	1982	Quest Books (IL)
48409	0345340833	Adrift: Seventy-Six Days Lost at Sea	Steven Callahan	1991	Ballantine Books
109817	0609604074	Trans-Sister Radio	Chris A. Bohjalian	2000	Harmony Books
240089	1862322929	A Cup Of Kindness: A History Of The Royal Scot...	Justine Taylor	2004	Tuckwell Press



Data Dictionary

Variável	Data Type	Descrição
User-ID	int	Unique identifier for the user (anonymized).
ISBN	str	International Standard Book Number, unique identifier for books.
Book-Rating	int	Rating given to a book by a user (0 = implicit, 1-10 = explicit).
Location	str	User's location in the format "city, state, country".
Age	float	User's age in years.
Book-Title	str	Title of the book.
Book-Author	str	Author of the book (first author if multiple).
Year-Of-Publication	int	Year the book was published.
Publisher	str	Name of the book's publisher.

Link do repositório: <https://github.com/gustavosrios/mack-semester-4>