

COMPONENTE CURRICULAR:	PROJETO APLICADO III
NOME COMPLETO DO ALUNO:	BRENDA LOUIZE DE O. SOUSA CABRAL – RA 10424949 CRISTINA ALMEIDA DA SILVA – RA 10424207 ÉLIDA ROSA DE PAIVA SOUZA – RA 10424468 ISABEL CABRAL VIEIRA DE SOUSA – RA 1042479

TÍTULO: SISTEMA DE RECOMENDAÇÃO DE LIVROS

fevereiro/2024

1. INTRODUÇÃO	3
2. MOTIVAÇÃO	4
3. JUSTIFICATIVA	5
4. OBJETIVOS	5
REFERÊNCIAS	7

1. INTRODUÇÃO

Com o crescimento exponencial das informações disponíveis na internet, identificar conteúdos relevantes de maneira eficiente tornou-se desafio significativo para os usuários. Nesse cenário, os sistemas de recomendação surgem como uma ferramenta essencial para personalizar a experiência do usuário, auxiliando na descoberta de livros compatíveis com os interesses individuais. Esses sistemas são amplamente utilizados em plataformas de streaming, e-commerce, redes sociais, bem como em plataformas de venda de livros online, bibliotecas digitais e serviços de leitura, impactando diretamente na experiência de leitura e na escolha dos usuários, com base em dados de comportamento e preferências anteriores (Resnick & Varian, 1997). Esse tipo de sistema não só facilita a descoberta de novos títulos, mas também oferece uma forma de otimizar a curadoria de conteúdos, tornando o processo mais ágil e personalizado (Ricci, Rokach & Shapira, 2015).

Para o desenvolvimento deste projeto, foi escolhida a base de dados **Book-Crossing Dataset**, disponibilizada por Ruchi798 na plataforma Kaggle (<https://www.kaggle.com/datasets/ruchi798/bookcrossing-dataset>). Esta base de dados foi originalmente coletada a partir do projeto Book-Crossing, criado por Cai-Nicolas Ziegler em 2004, e consiste em três conjuntos de dados inter-relacionados:

- **BX-Books.csv**: contém informações sobre os livros, incluindo ISBN, título, autor, ano de publicação, editora e URL da capa.
- **BX-Users.csv**: contém informações dos usuários, como ID, idade e localização.
- **BX-Book-Ratings.csv**: Registra as avaliações de livros feitas pelos usuários, podendo ser explícitas (de 1 a 10) ou implícitas (simplesmente indicando que um livro foi lido).

A base de dados cobre um período significativo de interações de leitura entre os usuários da plataforma Book-Crossing, proporcionando um amplo histórico para análise e modelagem de recomendações. Contudo, apresenta algumas limitações, como dados incompletos, inconsistências nos registros de idade dos usuários e distribuição desigual de avaliações entre os livros.

O pipeline de processamento de dados seguirá as seguintes etapas:

1. **Coleta e xarregamento dos dados**: Importação dos três conjuntos de dados (livros, usuários e avaliações) e verificação da qualidade dos dados.
2. **Limpeza e tratamento dos dados**: Remoção de valores inconsistentes, tratamento de valores ausentes e padronização de informações.

3. **Exploração e análise inicial:** Visualização e entendimento dos padrões dos dados para definir abordagens de modelagem.
4. **Construção do modelo de recomendação:** Implementação de técnicas de aprendizado de máquina, como filtragem colaborativa e baseada em conteúdo.
5. **Avaliação do modelo:** Medidas de desempenho, como RMSE (Root Mean Squared Error) e precisão das recomendações.
6. **Implementação e validação contínua:** Ajustes baseados no feedback dos usuários e aprimoramento contínuo do sistema.

A implementação desse pipeline permitirá a criação de um sistema de recomendação robusto, capaz de oferecer sugestões personalizadas e relevantes para os leitores, otimizando a experiência em plataformas digitais de leitura.

2. MOTIVAÇÃO

A intensa demanda por soluções personalizadas, sobretudo em um ambiente saturado de informações como a internet, tem impulsionado o desenvolvimento de sistemas de recomendação mais eficazes. Segundo *Resnick & Varian* (1997), os sistemas de recomendação desempenham um papel crucial ao ajudar os usuários a filtrar a vasta quantidade de opções disponíveis, promovendo uma experiência mais direcionada e personalizada. Neste contexto, o uso de técnicas de aprendizado de máquina, que lidam com grandes volumes de dados e buscam melhorar a precisão das recomendações, torna-se essencial. *Aggarwal* (2016) destaca que algoritmos de aprendizado de máquina podem não apenas melhorar a relevância das sugestões, mas também lidar com a dinâmica e a complexidade dos dados comportamentais dos usuários.

A escolha de focar em um sistema de recomendação de livros é motivada pela necessidade de aprimorar a experiência do usuário, oferecendo sugestões personalizadas e ajustadas aos seus gostos e comportamentos de leitura. Como afirmam *Ricci, Rokach & Shapira* (2015), a personalização é fundamental para melhorar a satisfação do usuário, tornando a interação com plataformas de conteúdo mais relevante e engajante. Este projeto também proporciona uma oportunidade de aplicar teorias e conceitos aprendidos ao longo da graduação em Ciência de Dados, enfrentando um desafio prático e relevante, como ressaltado por *Aggarwal* (2016), que destaca a aplicabilidade de métodos preditivos para resolver problemas concretos em diversas indústrias.

3. JUSTIFICATIVA

A escolha de um sistema de recomendação de livros é justificada pela sua relevância crescente no mercado editorial digital, que tem visto uma transição de vendas físicas para plataformas digitais. De acordo com *Jannach et al.* (2010), a personalização em sistemas de recomendação tem se mostrado uma ferramenta poderosa para melhorar a experiência do usuário em plataformas digitais, oferecendo não apenas maior satisfação, mas também fidelização do público. Com uma quantidade crescente de títulos e resenhas disponíveis online, o problema de ajudar os leitores a selecionar livros relevantes se torna cada vez mais urgente, como apontado por *Basilico & Ricci* (2004), que discutem a importância de personalizar as recomendações para garantir que o usuário encontre facilmente o conteúdo que mais lhe interessa.

A utilização de técnicas de aprendizado de máquina para otimizar modelos de recomendação não só torna esse processo mais eficiente, como também garante uma experiência mais fluida e satisfatória para o usuário. Como argumenta *Schafer et al.* (2007), a aplicação de modelos avançados de recomendação, como a filtragem colaborativa e baseada em conteúdo, pode oferecer resultados mais precisos e adaptativos, proporcionando uma experiência personalizada e ajustada às preferências do usuário. Este projeto representa não apenas uma aplicação prática dos conceitos de aprendizado de máquina, mas também uma oportunidade para desenvolver habilidades cruciais de análise de dados e programação, essenciais para a formação de um cientista de dados.

A justificativa também se apoia no impacto direto que um sistema de recomendação de livros pode ter em plataformas de leitura, livrarias online e bibliotecas digitais. *Ricci et al.* (2015) afirmam que os sistemas de recomendação podem ajudar a aumentar o alcance de livros menos conhecidos e promover novos lançamentos, proporcionando maior diversidade de conteúdo e ampliando a visibilidade de obras que poderiam passar despercebidas. Essa inovação na experiência do usuário é um ganho tanto para os leitores quanto para as plataformas que utilizam o sistema, o que reforça sua importância na indústria editorial atual.

4. OBJETIVOS

O objetivo principal deste projeto é desenvolver um sistema de recomendação de livros utilizando técnicas de aprendizado de máquina, tendo como base o Book-Crossing Dataset, disponibilizado no Kaggle por Ruchi Bhatia. Esse conjunto de dados contém

informações sobre livros, usuários e avaliações, permitindo a construção de modelos personalizados para sugerir títulos alinhados aos interesses individuais dos leitores. O sistema buscará otimizar a descoberta de novos livros a partir das preferências, avaliações e histórico de leitura dos usuários, melhorando a experiência de navegação em plataformas digitais de leitura e aumentando a satisfação do usuário.

Objetivos Específicos

- **Explorar e compreender o dataset Book-Crossing:** realizar uma análise detalhada dos três conjuntos de dados disponibilizados (livros, usuários e avaliações), identificando sua estrutura, padrões e possíveis limitações.
- **Realizar limpeza e tratamento dos dados:** identificar e lidar com valores ausentes, inconsistências e possíveis ruídos nos dados para garantir a qualidade da modelagem.
- **Aplicar técnicas de aprendizado de máquina:** utilizar algoritmos como filtragem colaborativa, filtragem baseada em conteúdo e técnicas híbridas para construir modelos preditivos capazes de recomendar livros com base no perfil e nas preferências do usuário.
- **Avaliar o desempenho dos modelos:** empregar métricas como RMSE (Root Mean Squared Error), precisão e recall para medir a qualidade das recomendações geradas.
- **Desenvolver e testar o sistema de recomendação:** implementar um protótipo funcional baseado nos dados do Book-Crossing Dataset, verificando sua eficácia na geração de recomendações personalizadas.
- **Aprimorar a personalização contínua:** incorporar funcionalidades que permitam aos usuários fornecer feedback sobre as recomendações, possibilitando a adaptação e a evolução contínua das sugestões com base nas preferências individuais e nos comportamentos de leitura.
- **Desenvolver habilidades técnicas e analíticas:** aplicar os conhecimentos adquiridos durante o curso de Ciência de Dados, desenvolvendo habilidades práticas e técnicas em aprendizado de máquina, análise de dados e implementação de sistemas de recomendação, com ênfase em uma aplicação prática e realista no contexto editorial digital.

REFERÊNCIAS

- AGGARWAL, Charu C. *Recommender Systems: The Textbook*. Springer, 2016. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-29659-3>. Acesso em: 1 mar. 2025.
- BASILICO, Justin; RICCI, Francesco. *Adaptive Recommender Systems: An Experimental Evaluation*. In: Proceedings of the 5th ACM Conference on Electronic Commerce, 2004, p. 239-246.
- JANACH, D.; ADOMAVICIUS, G. *Recommender Systems: Challenges and Research Opportunities*. Computer Science Review, 2010. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1574013710000147>. Acesso em: 1 mar. 2025.
- RESNICK, P.; VARIAN, H. R. *Recommender Systems*. Communications of the ACM, v. 40, n. 3, p. 56-58, 1997. Disponível em: <https://dl.acm.org/doi/10.1145/245108.245121>. Acesso em: 1 mar. 2025.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. *Recommender Systems Handbook*. 2. ed. Springer, 2015. Disponível em: <https://link.springer.com/book/10.1007/978-1-4899-7637-6>. Acesso em: 1 mar. 2025.
- RUCHI, B. Conjunto de dados de cruzamento de livros. 2022. Disponível em: <https://www.kaggle.com/datasets/ruchi798/bookcrossing-dataset>. Acesso em: 1 mar. 2025.
- SCHAFER, J. Ben; KONSTAN, Joseph A.; RIEDL, *Recommender Systems: Challenges and Opportunities*. Computer Science Review, 2007. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1574013707000096>. Acesso em: 1 mar. 2025.
- SCHAFER, J. Ben; KONSTAN, Joseph A.; RIEDL, John. *E-commerce recommendation applications*. *Data Mining and Knowledge Discovery*, v. 5, p. 115-153, 2001.
- ZIEGLER, Cai-Nicolas. *Book-Crossing Dataset*. Book-Crossing, 2004. Disponível em: <http://www.informatik.uni-freiburg.de/~cziegler/BX/>. Acesso em: 2 mar. 2025.