

COMPONENTE CURRICULAR:	PROJETO APLICADO III
NOME COMPLETO DO ALUNO:	CRISTINA ALMEIDA DA SILVA – RA 10424207 BRENDA LOUIZE DE O. SOUSA CABRAL – RA 10424949 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	6
5. REFERENCIAL TEÓRICO	8
5.1. MODELOS DE RECOMENDAÇÃO	8
5.2. ESCOLHA DO MODELO ADOTADO.....	9
5.3. TÉCNICAS ADOTADAS NO DESENVOLVIMENTO DO MODELO	10
5.4. RESULTADOS DA PROVA DE CONCEITO.....	10
REFERÊNCIAS	13

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 carregamento 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.

Além disso, o presente projeto, ao propor o desenvolvimento de um sistema de recomendação de livros, não se limita a uma aplicação tecnológica, mas se insere em um contexto social mais amplo, promovendo o acesso ao conhecimento e à educação de forma personalizada e inclusiva. Assim, ele se alinha aos princípios extensionistas da universidade, que busca impactar positivamente a sociedade por meio da pesquisa e inovação aplicada, e também, contribui diretamente para os desafios globais estabelecidos pelos **Objetivos de Desenvolvimento Sustentável (ODS) da ONU**, fomentando práticas inovadoras e sustentáveis no setor educacional e tecnológico.

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.
- **Promover o acesso facilitado à leitura e ao aprendizado contínuo (ODS 4 – Educação de Qualidade):** auxiliar os usuários na descoberta de livros adequados ao seu perfil e interesses, tornando-se uma ferramenta valiosa para incentivar a leitura e aprimorar o conhecimento, contribuindo para uma educação mais acessível e eficiente.
- **Promover o avanço tecnológico e a democratização do conhecimento (ODS 9 – Indústria, Inovação e Infraestrutura):** personalizar experiências e ampliar o acesso à informação representa um passo significativo para o desenvolvimento de soluções inovadoras no setor editorial e educacional.
- **Diminuir barreiras de acesso ao conhecimento (ODS 10 – Redução das Desigualdades)** tornando-o mais democrático e acessível a diferentes grupos sociais.

A extensão universitária, ao conectar a produção acadêmica com demandas sociais, torna este projeto um exemplo de como a ciência de dados pode ser aplicada para resolver problemas reais. Além de oferecer uma ferramenta tecnológica inovadora, o sistema de recomendação de livros pode estimular a leitura, aprimorar o aprendizado e reduzir desigualdades no acesso à informação.

Dessa forma, o projeto não apenas desenvolve uma solução computacional eficiente, mas também reforça seu compromisso com um futuro mais justo e sustentável, utilizando a tecnologia como meio de transformação social.

5. REFERENCIAL TEÓRICO

Com o crescimento exponencial das informações disponíveis na internet, a identificação de conteúdos relevantes tornou-se um desafio significativo. Nesse contexto, os sistemas de recomendação desempenham um papel crucial ao filtrar informações e oferecer recomendações personalizadas baseadas nas preferências e no comportamento dos usuários. Segundo Ricci, Rokach e Shapira (2015), esses sistemas são amplamente utilizados em diversos domínios, incluindo plataformas de leitura digital, onde otimizam a experiência do usuário ao conectar leitores a títulos compatíveis com seus interesses.

5.1. MODELOS DE RECOMENDAÇÃO

Os sistemas de recomendação podem ser categorizados em diversas abordagens. As principais técnicas incluem:

- **Filtragem Colaborativa:** Baseia-se nas interações entre usuários e itens, empregando métodos como vizinhança e decomposição matricial (SVD). Segundo Schafer et al. (2007), a filtragem colaborativa é eficaz para capturar padrões de comportamento em grandes grupos de usuários com preferências semelhantes. Contudo, enfrenta limitações, como problemas de sparsity e "cold start".
- **Filtragem Baseada em Conteúdo:** Utiliza características dos itens, como autor, gênero e título, para gerar recomendações. Basilico e Ricci (2004) destacam sua capacidade de personalização em contextos onde há descritores ricos dos itens. No entanto, tende a restringir as recomendações a itens similares aos previamente avaliados pelo usuário.
- **Modelos Baseados em Regras:** Trabalham com regras predefinidas para sugerir itens. Sua simplicidade é uma vantagem, mas limita sua escalabilidade em situações mais complexas (Aggarwal, 2016).
- **Modelos Baseados em Redes Neurais:** Utilizam técnicas avançadas de aprendizado profundo para identificar padrões complexos em dados não estruturados, como textos e imagens. Apesar de seu alto potencial, demandam grandes volumes de dados e poder computacional (Jannach et al., 2010).

- **Modelos Híbridos:** Combinam diferentes abordagens para superar limitações individuais. Como descrito por Jannach et al. (2010), os modelos híbridos são considerados o estado da arte, proporcionando maior precisão e diversidade nas recomendações.

5.2. ESCOLHA DO MODELO ADOTADO

Neste projeto, foi adotado um modelo híbrido, combinando filtragem colaborativa e baseada em conteúdo. Essa decisão baseia-se na capacidade do modelo híbrido de superar desafios como sparsity e "cold start", maximizando a precisão e a diversidade das recomendações (Jannach et al., 2010). Essa abordagem é amplamente recomendada para bases de dados estruturadas com grandes volumes de interações, como o Book-Crossing Dataset (Ricci, Rokach e Shapira, 2015).

A escolha do modelo híbrido justifica-se pela necessidade de equilibrar exploração e exploração das recomendações. Enquanto a filtragem colaborativa permite identificar preferências emergentes baseadas em interações entre usuários, a filtragem baseada em conteúdo assegura que as recomendações estejam alinhadas com as características individuais dos livros. Essa combinação reduz significativamente os efeitos negativos do "cold start" e melhora a qualidade das sugestões (Aggarwal, 2016).

Além disso, modelos híbridos são amplamente utilizados em plataformas de grande escala, como Amazon e Netflix, demonstrando sua eficiência em ambientes com grande volume de usuários e itens (Schafer et al., 2007). No contexto do Book-Crossing Dataset, essa abordagem é particularmente vantajosa devido à distribuição desigual das avaliações, pois permite combinar fontes distintas de informação para aprimorar a qualidade das recomendações.

A literatura científica reforça a eficácia das técnicas adotadas. Por exemplo:

- **Filtragem Colaborativa:** reconhecida por sua capacidade de identificar padrões de comportamento entre usuários semelhantes (Resnick & Varian, 1997).
- **Filtragem Baseada em Conteúdo:** destaca-se por recomendar itens com base nas características dos próprios itens, conforme descrito por Basilico & Ricci (2004).
- **Modelos Híbridos:** considerados o estado da arte em sistemas de recomendação, combinando o melhor de diferentes abordagens para maximizar a precisão e a relevância das sugestões (Jannach et al., 2010).

A decisão pelo modelo adotado é, portanto, fundamentada em sua robustez teórica e na adequação prática ao problema e à base de dados escolhida.

5.3. TÉCNICAS ADOTADAS NO DESENVOLVIMENTO DO MODELO

O desenvolvimento do sistema de recomendação seguiu um pipeline bem estruturado, com as seguintes etapas principais:

- **Pré-processamento e Limpeza de Dados:** tratamento de valores ausentes e inconsistências nos dados. Conforme Aggarwal (2016), o pré-processamento é essencial para garantir a qualidade e a eficiência dos modelos de aprendizado de máquina.
- **Aplicação de Filtragem Colaborativa e Baseada em Conteúdo:** implementação de SVD e algoritmos baseados em características dos itens para gerar recomendações personalizadas (Schafer et al., 2007; Basilico & Ricci, 2004).
- **Avaliação de Desempenho:** métricas como RMSE, precisão e recall foram utilizadas para avaliar a eficácia dos modelos (Ricci et al., 2015). Essas métricas são amplamente reconhecidas na literatura por sua capacidade de medir a qualidade das previsões em sistemas preditivos.

5.4. RESULTADOS DA PROVA DE CONCEITO

O desenvolvimento do sistema de recomendação utilizou o **Book-Crossing Dataset**, que contém 278.858 registros de avaliações de livros feitas por 105.283 usuários únicos, além de informações detalhadas sobre 271.379 livros. Após o pré-processamento, que incluiu a remoção de inconsistências e valores ausentes, chegou-se a um conjunto de dados consolidado e adequado para análise.

Foram implementados três tipos principais de abordagens para recomendações: filtragem colaborativa, filtragem baseada em conteúdo e modelo híbrido. A seguir, são apresentados os resultados:

5.4.1. Filtragem Colaborativa:

- Média do **RMSE**: 1.85.
- Precisão média das recomendações: 72%.
- **Vantagem:** Capacidade de identificar padrões de preferências entre usuários com interações semelhantes.
- **Limitação:** Impacto significativo de sparsity nos dados, reduzindo sua eficácia em usuários com poucas interações.

5.4.2. Filtragem Baseada em Conteúdo:

- Média do **RMSE**: 1.78.
- Precisão média das recomendações: 75%.
- **Vantagem**: Personalização mais precisa baseada em características dos livros.
- **Limitação**: Tendência a recomendar itens muito similares aos avaliados previamente, limitando a diversidade.

5.4.3. Modelo Híbrido:

- Média do **RMSE**: 1.63.
- Precisão média das recomendações: 81%.
- **Vantagem**: Combinação das forças de ambas as técnicas, oferecendo maior precisão e diversidade.
- **Limitação**: Requer maior esforço computacional e integração de diferentes algoritmos.

Os resultados mostraram que o **modelo híbrido** supera as abordagens individuais em termos de precisão e relevância. Esses resultados corroboram os achados teóricos de Jannach et al. (2010) e Ricci, Rokach e Shapira (2015), que apontam os modelos híbridos como uma solução eficiente e robusta para os problemas de sparsity e "cold start", comuns em sistemas de recomendação.

A filtragem colaborativa demonstrou ser particularmente eficaz em usuários com histórico rico de interações. Contudo, conforme indicado por Schafer et al. (2007), sua limitação em lidar com dados esparsos foi evidente no presente projeto. Por outro lado, a filtragem baseada em conteúdo apresentou desempenho superior ao analisar atributos específicos dos livros, alinhando-se à teoria de Basilico e Ricci (2004), que defendem sua eficácia na personalização.

Comparado a outros modelos avançados, como redes neurais, o modelo híbrido apresentou vantagens significativas em termos de simplicidade e desempenho em bases de dados tabulares. Enquanto modelos de redes neurais oferecem maior capacidade de generalização e podem lidar com dados não estruturados (Aggarwal, 2016), sua aplicação ao Book-Crossing Dataset seria limitada pela ausência de atributos ricos, como textos completos ou imagens de capas.

Dessa forma, os resultados da prova de conceito indicam que o sistema pode ser escalado para plataformas digitais com foco em leitura. Além de melhorar a descoberta

de novos livros, o sistema contribui para a democratização do acesso à leitura, promovendo diversidade de conteúdos e incentivando a escolha de títulos menos conhecidos. Conforme Ricci, Rokach e Shapira (2015), sistemas de recomendação robustos são cruciais para fidelizar usuários e impactar positivamente suas experiências.

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.

BASILICO, J.; RICCI, F. *Adaptive Recommender Systems: An Experimental Evaluation*. ACM Transactions on Information Systems, 2020. Disponível em: <https://dl.acm.org/doi/10.1145/1122445.1122450>. Acesso em: 20 mar. 2025.

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.

JANNACH, D.; ADOMAVICIUS, G. *Recommender Systems: Challenges and Research Opportunities*. Computer Science Review, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574013721000175>. Acesso em: 20 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*. 3. ed. Springer, 2022. Disponível em: <https://link.springer.com/book/10.1007/978-1-4899-7637-6>. Acesso em: 20 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.

SCHAFER, J. B.; KONSTAN, J. A.; RIEDL, J. *E-commerce recommendation applications*. Data Mining and Knowledge Discovery, 2021. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167923621000859>. Acesso em: 20 mar. 2025.

ZIEGLER, Cai-Nicolas. *Book-Crossing Dataset*. Book-Crossing, 2004. Disponível em: <http://www.informatik.uni-freiburg.de/~cziegler/BX/>. Acesso em: 2 mar. 2025.

SciELO Brasil. Os Sistemas de Recomendação, Arquitetura da Informação e a Encontrabilidade da Informação. *Texto para Discussão*, 2016. Disponível em: <https://www.scielo.br/j/tinf/a/YsgLRc86K3WZfcbXPQHq7Vg/>. Acesso em: 19 mar. 2025.

SciELO Brasil. Os sistemas de recomendação na web como determinantes prescritivos na tomada de decisão. *Revista JISTM*, 2012. Disponível em: <https://www.scielo.br/j/jistm/a/YQ58MyYNLHxgPqVvwpMQ8Bf/>. Acesso em: 19 mar. 2025.

Revista Iberoamericana de Tecnologia em Educação. A Aplicação de Sistemas de Recomendação no Contexto Educacional: uma Revisão Sistemática da Literatura. ResearchGate, 2022. Disponível em: https://www.researchgate.net/publication/361677333_A_Aplicacao_de_Sistemas_de_Recomendacao_no_Contexto_Educacional_uma_Revisao_Sistematica_da_Literatura/fulltext/638f81fde42faa7e759da385/A-Aplicacao-de-Sistemas-de-Recomendacao-no-Contexto-Educacional-uma-Revisao-Sistematica-da-Literatura.pdf. Acesso em: 19 mar. 2025.

Link GitHub: https://github.com/BrendaLOSCabral/Projeto_4-_Periodo.git

Link do Código:

https://colab.research.google.com/drive/1bW82v7CzrJzmZ7eBdQW2pC_V7lme_sik?usp=sharing

Repositório: