



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 Brasília**

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 | Referencial Teórico..... |
| 4 | Metodologia..... |
| 5 | Referências |
| 5 | |



Introdução

No contexto atual, em que as tecnologias digitais moldam grande parte das experiências cotidianas, plataformas como YouTube, Amazon e Netflix popularizaram-se por meio de sistemas de recomendação que personalizam o consumo de conteúdo.

A partir dessa observação, surge o questionamento sobre como essa abordagem pode ser aplicada para incentivar o hábito de leitura. O presente projeto propõe investigar como os sistemas de recomendação de livros podem contribuir para o aumento do interesse pela leitura, especialmente em um cenário onde a oferta de entretenimento digital tem predominado.

A leitura é uma atividade que oferece benefícios profundos e duradouros, tanto para o desenvolvimento pessoal quanto para a formação de uma sociedade mais crítica e informada. No entanto, com o tempo cada vez mais escasso e a vasta oferta de livros, a escolha de uma obra que realmente atraia o leitor pode se tornar um desafio.

A leitura tem o potencial de transformar vidas, oferecendo novas perspectivas, conhecimentos e entretenimento. No entanto, com a crescente popularidade de outras formas de mídia, o hábito de leitura muitas vezes acaba ficando em segundo plano. Neste contexto, sistemas de recomendação de livros podem desempenhar um papel fundamental ao ajudar leitores a encontrar obras que correspondam aos seus gostos e interesses, incentivando-os a ler mais e manter o hábito ao longo da vida.

Nesse sentido, o desenvolvimento de um sistema de recomendação eficaz, que utilize dados demográficos e comportamentais para sugerir leituras personalizadas, pode ser uma solução relevante para promover a leitura de forma mais acessível e satisfatória.

Este projeto surge da constatação de que muitos leitores, após se encantarem com um livro, não sabem quais obras buscar em seguida, o que pode desestimular a continuidade do hábito de leitura. A proposta deste estudo é, portanto, desenvolver um sistema de recomendação que conecte leitores a livros semelhantes aos que já leram, utilizando algoritmos de inteligência artificial. O objetivo é criar uma solução que ajude os leitores a descobrir novos títulos de forma personalizada, promovendo a leitura contínua e enriquecendo a experiência literária.



Motivação:

- A motivação central deste trabalho é estimular o hábito de leitura, utilizando tecnologias de ciência de dados para oferecer recomendações personalizadas de livros. Em um ambiente onde a concorrência com outras formas de entretenimento é intensa, acredita-se que uma abordagem direcionada e personalizada pode facilitar o acesso dos leitores a obras que atendam suas preferências, resultando em uma experiência literária mais gratificante.

Justificativa:

- O incentivo à leitura possui um papel fundamental na formação de indivíduos mais críticos e informados, sendo uma ferramenta poderosa de transformação social. A dificuldade em encontrar livros que correspondam aos interesses pessoais pode desmotivar muitos leitores, especialmente aqueles que não possuem um hábito consolidado de leitura. A implementação de um sistema de recomendação de livros, que leve em consideração tanto preferências individuais quanto dados demográficos, é uma proposta relevante para fortalecer o hábito de leitura e aumentar o acesso a conteúdos literários de qualidade, promovendo a inclusão e a democratização da cultura.

Objetivo Geral:

- Desenvolver um sistema de recomendação de livros que combine métodos de filtragem colaborativa e baseados em conteúdo, com foco em personalização e relevância, visando a promoção da leitura e a valorização do hábito literário.

Objetivos Específicos:

- Avaliar o impacto de fatores demográficos, como idade e localização, nas preferências de leitura.
- Analisar a qualidade dos dados disponíveis e seus efeitos na precisão das recomendações.
- Implementar e testar diferentes algoritmos de recomendação, comparando seu desempenho com base em métricas de avaliação.
- Desenvolver um sistema híbrido de recomendação, que integre filtragem colaborativa e métodos baseados em conteúdo, garantindo sugestões mais relevantes e personalizadas.
- Explorar o impacto da personalização nas recomendações e sua influência na satisfação dos leitores e no estímulo à leitura.

"Um livro não é apenas um amigo, ele faz amigos 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)



Referencial Teórico

O desenvolvimento de sistemas de recomendação é amplamente fundamentado na literatura de recuperação de informações e aprendizado de máquina. Esses sistemas são ferramentas importantes para filtrar e sugerir itens relevantes com base em características específicas dos usuários ou dos itens.

Um dos métodos mais utilizados é a Filtragem Baseada em Conteúdo (*Content-Based Filtering*), que se concentra na similaridade entre os itens recomendados e os que já foram consumidos pelo usuário. Essa abordagem tem suas raízes em teorias de recuperação da informação, onde a semelhança entre documentos é avaliada de acordo com suas características textuais ou metadados.

Estudos de *Pazzani* e *Billsus* (2007) são referências importantes nesse campo, pois descrevem as melhores práticas e os desafios encontrados na construção de sistemas de recomendação baseados em conteúdo. A avaliação da similaridade entre itens é feita utilizando descritores textuais, como título, descrição, autores e categorias. Assim, o sistema recomenda novos livros com atributos semelhantes aos já avaliados positivamente pelo usuário (*PAZZANI; BILLSUS, 2007*).

A transformação de texto em vetores numéricos, necessária para a avaliação de similaridade, é realizada utilizando a técnica de *TF-IDF* (*Term Frequency-Inverse Document Frequency*). O método *TF-IDF*, descrito por Salton e Buckley (1988), permite medir a relevância de cada termo em um documento, considerando tanto a frequência do termo dentro do documento quanto sua raridade no conjunto total de dados. Isso garante que termos mais informativos tenham maior peso no cálculo de similaridade (*SALTON; BUCKLEY, 1988; RAMOS, 2003*).

Para calcular a similaridade entre os itens, foi utilizada a Similaridade de Cosseno (*Cosine Similarity*). Essa métrica avalia o ângulo entre os vetores *TF-IDF* de dois itens, onde um valor de similaridade de cosseno mais próximo de 1 indica uma maior semelhança entre os itens (*SINGHAL, 2001*). A Similaridade de Cosseno é uma técnica eficiente e amplamente adotada para mensurar similaridades em sistemas de recomendação baseados em conteúdo.

O embasamento teórico das técnicas utilizadas, como a filtragem baseada em conteúdo, *TF-IDF* e Similaridade de Cosseno, não só valida a escolha das metodologias aplicadas, mas também orienta a implementação e avaliação do sistema de recomendação. Além disso, práticas como o pré-processamento de dados textuais, incluindo a limpeza e a normalização, garantem a qualidade dos dados, que é essencial para o desempenho do modelo.



Metodologia





1. Os dados foram obtidos pelo Kaggle -

<https://www.kaggle.com/dylanjcastillo/7k-books-with-metadata>.

2. Análises iniciais foram feitas como verificações de integridade dos dados, tipagem e estatísticas descritivas.
3. Foi realizado o tratamento de outliers, dados nulos e duplicados. Novas features foram criadas para a próxima etapa de análise exploratória.
4. Na EDA foram feitas análises de correlação com heatmaps, análises de distribuição com histogramas.
5. No pré-processamento as variáveis de texto '*description*', '*authors*' e '*categories*' foram tratadas e combinadas em uma só variável '*combined_features*' que por sua vez foi vetorizada pelo *TfidfVectorizer* atribuindo um valor numérico as palavras com base em suas ocorrências.
6. O algoritmo foi modelado com base na Similaridade entre Cossenos (em inglês *Cosine Similarity*), o sistema de recomendação personalizado recebe os títulos dos livros e retorna recomendações de livros semelhantes aos enviados.
7. Na avaliação do modelo foram feitos testes em diversas amostras de livros selecionadas aleatoriamente, e o algoritmo retornou corretamente os títulos mais semelhantes.



Referências

PAZZANI, M.; BILLSUS, D. *Content-based recommendation systems*. In: The Adaptive Web. Springer, 2007. p. 325-341.

RAMOS, J. *Using TF-IDF to determine word relevance in document queries*. In: *Proceedings of the first instructional conference on machine learning*. 2003.

SALTON, G.; BUCKLEY, C. *Term-weighting approaches in automatic text retrieval*. *Information Processing & Management*, v. 24, n. 5, p. 513-523, 1988.

SINGHAL, A. *Modern information retrieval: A brief overview*. *IEEE Data Engineering Bulletin*, v. 24, n. 4, p. 35-43, 2001.

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