



UNIVERSIDADE PRESBITERIANA MACKENZIE

TECNÓLOGO EM CIÊNCIAS DE DADOS

Caroline Ribeiro Ferreira – 10408052

Lais César Fonseca – 10407066

Leonardo dos Reis Olher – 10407752

Liliane Gonçalves de Brito Ferraz – 10407087

Múcio Emanuel Feitosa Ferraz Filho – 10218691

Otávio Bernardo Scandiuzzi – 10407867

SISTEMA DE RECOMENDAÇÃO DE LIVROS

SÃO PAULO

2024



Sumário

1.	GLOSSÁRIO	3
2.	RESUMO	4
3.	INTRODUÇÃO	5
4.	OBJETIVO.....	7
5.	CRONOGRAMA.....	8
6.	METODOLOGIA	9
7.	RESULTADOS	10
8.	CONCLUSÃO	11
9.	DIRETÓRIO GITHUB	12
10.	REFERENCIAL TEÓRICO	13



1. GLOSSÁRIO

- **Colaboratory:** Conhecido também como “Colab”, é um produto do Google Research, área de pesquisas científicas do Google. O Colab permite que qualquer pessoa escreva e execute código Python arbitrário pelo navegador e é especialmente adequado para aprendizado de máquina, análise de dados e educação.
- **DataFrame:** É uma estrutura de dados bidimensional com os dados alinhados de forma tabular em linhas e colunas.
- **Datasets:** conjuntos de dados organizados em um formato similar ao das tabelas, com linhas e colunas que contém informações sobre determinado tema.
- **GitHub:** GitHub é uma plataforma de hospedagem de código-fonte e arquivos com controle de versão usando o Git. Ele permite que programadores, utilitários ou qualquer usuário cadastrado na plataforma contribuam em projetos privados e/ou Open Source de qualquer lugar do mundo.
- **Kaggle:** É uma plataforma para aprendizado de ciência de dados. É também uma comunidade, a maior da internet, para assuntos relacionados com Data Science.
- **Machine Learning (Aprendizado de máquina):** É um subcampo da Engenharia e da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial.
- **Python:** É uma linguagem de programação de alto nível, interpretada de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991.
- **Streaming:** Fluxo contínuo, fluxo de média, fluxo de mídia ou transmissão contínua, é uma forma de distribuição digital, em oposição à descarga de dados.



2. RESUMO

Neste projeto, propomos o desenvolvimento de um sistema de recomendação de livros utilizando técnicas de aprendizado de máquina e análise estatística preditiva. O objetivo principal é melhorar a experiência do usuário ao fornecer recomendações personalizadas individual com base em seu histórico de leituras.

Para o desenvolvimento do projeto utilizaremos um dataset público, que possuem acesso livre no site Kaggle criando um modelo de recomendação de livros, para recomendar livros semelhantes ao leitor com base em seu interesse. Para isto, utilizaremos dos métodos adquiridos nos componentes curriculares de introdução a ciência de dados, como aplicação da linguagem Python com o uso do Colab.

A abordagem do projeto visa aplicarmos o método de filtragem baseada em colaboração, no qual precisamos construir uma máquina preditiva que, com base nas escolhas de leituras de outras pessoas, o livro seja recomendado a outras pessoas com interesses semelhantes.

3. INTRODUÇÃO

No atual cenário literário, a vasta gama de opções de livros disponíveis desafia os leitores a navegarem por um oceano de informações para encontrar obras que verdadeiramente cativem e ressoem com seus interesses individuais. Diante desse desafio, surge a motivação para o desenvolvimento deste projeto: um “Sistema de Recomendações de Livros”. Esta introdução abordará o contexto do trabalho, a motivação, a justificativa e os objetivos, proporcionando uma visão holística do propósito e das metas deste projeto aplicado.

De acordo com o portal de notícias G1, pertencente ao grupo Globo, apenas 16% da população brasileira comprou algum livro no ano de 2023, verifica-se dessa forma a necessidade de impulsionar um mercado tão importante para o desenvolvimento intelectual das pessoas. Um possível reflexo disso foi o baixo desempenho dos jovens brasileiros no Ranking PISA, acendendo um sinal de alerta para esta situação. Desta forma, buscamos encontrar meios que facilitem o acesso à leitura, de forma que seja otimizada a escolha de bons livros que possam impulsionar o desenvolvimento de hábitos saudáveis naqueles que o utilizarem.

No contexto atual, a abundância de opções literárias é tanto um privilégio quanto um desafio. A tecnologia, no entanto, oferece oportunidades significativas para aprimorar a experiência de leitura, simplificando a escolha de livros por meio de sistemas inteligentes de recomendação. Estes, ao analisarem padrões de leitura, hábitos e preferências individuais, podem fornecer sugestões personalizadas, transformando a busca por livros em uma jornada mais enriquecedora.

A motivação para este projeto emerge da percepção das dificuldades enfrentadas pelos leitores contemporâneos ao selecionar livros entre uma ampla variedade de opções. A escolha deste tema é impulsionada pelo desejo de simplificar e aprimorar a experiência de leitura, tornando-a mais personalizada e acessível. Além disso, a paixão pela literatura e o reconhecimento do impacto positivo que uma recomendação bem-sucedida pode ter na vida do leitor constituem fortes motivadores.

Justificamos o desenvolvimento deste sistema pela sua capacidade potencial de democratizar o acesso à literatura, ampliando o alcance de obras significativas e diversificando o repertório literário dos usuários. Acreditamos que o projeto pode não apenas facilitar a escolha de livros, mas também promover a descoberta de novos gêneros e autores, enriquecendo assim a experiência de leitura de forma significativa.

Os sistemas de recomendação de livros desempenham um papel crucial não só na facilitação da descoberta de novas obras para os leitores, mas também na promoção da diversidade



cultural, no estímulo ao diálogo intercultural e na ampliação do conhecimento sobre uma vasta gama de culturas e perspectivas, com isto atendemos as questões relacionadas aos ODS (Objetivos de Desenvolvimento Sustentável no Brasil).

Ao sugerir livros que se relacionam com interesses específicos ou áreas de estudo, esses sistemas auxiliam os usuários a enriquecer seu entendimento sobre diferentes culturas, períodos históricos e pontos de vista. Além disso, ao possibilitar a descoberta de livros de uma variedade de gêneros, autores e temas, os sistemas de recomendação contribuem para enriquecer a diversidade cultural presente na leitura dos usuários.

A abordagem híbrida foi a escolhida para compor este projeto, aliando a filtragem colaborativa, baseada em interações de usuários, e a filtragem baseada em conteúdo, que foca nas características dos livros presentes nos arquivos. O uso da aprendizagem supervisionada auxiliará na obtenção de informações de múltiplos tipos de dados, como dados textuais, avaliações de usuários e demais metadados editoriais, todos esses divididos entre três datasets escolhidos para compor este projeto. Isso resultará em um modelo final de machine learning que pode prever diretamente os interesses dos usuários com a adição de itens específicos.



4. OBJETIVO

Este projeto tem como principal objetivo desenvolver um Sistema de Recomendações de Livros eficiente e inovador, aplicando técnicas avançadas de recomendação para proporcionar sugestões personalizadas aos usuários. Além disso, busca-se:

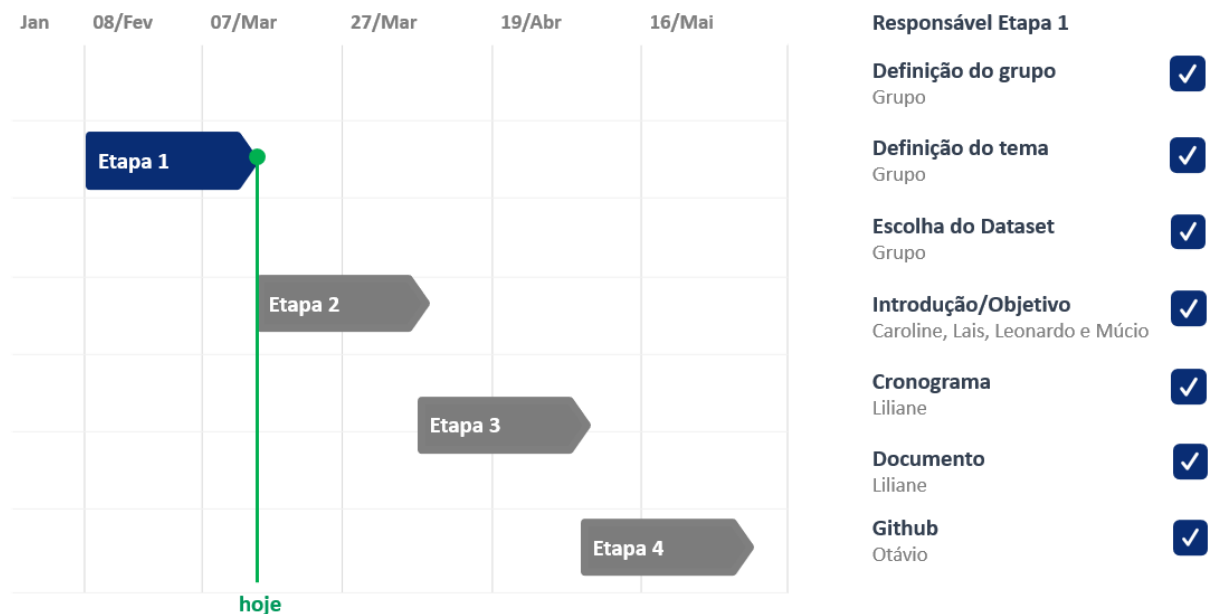
- Construir uma plataforma interativa que permita aos usuários interagirem entre si, compartilhando recomendações, resenhas e insights literários.
- Analisar e avaliar criticamente diferentes abordagens de recomendação, propondo melhorias e ajustes conforme necessário.
- Aumentar o engajamento dos usuários e estimular o hábito da leitura, com o intuito de contribuir para a formação de uma comunidade de leitores mais diversificada, tornando a leitura mais inclusiva e abrangente, contribuir para uma sociedade mais informada, empática e culturalmente enriquecida.

Através desses objetivos, almejamos não apenas criar um sistema eficaz de recomendação de livros, mas também fomentar uma comunidade engajada e apaixonada pela leitura. Esses elementos fundamentais formam a base do nosso projeto, delineando a trajetória que percorreremos para alcançar nossos objetivos e oferecer uma solução valiosa no universo literário contemporâneo.



5. CRONOGRAMA

Reportar o desenvolvimento das etapas propostas e o percentual de evolução de entregas referente ao componente curricular de Projeto Aplicado III do curso de Tecnologia em Ciências de Dados.





6. METODOLOGIA



7. RESULTADOS



8. CONCLUSÃO



9. DIRETÓRIO GITHUB

Todo o conteúdo do projeto estará disponível no site da GitHub, que poderá ser acessado pelo link:

<https://github.com/OtavioBer/ProjAplicadoIII>

O diretório está organizado por pastas:

Pasta “Códigos” será disponibilizado os códigos em Python que foram utilizados para realizar a análise exploratória, tratamento dos dados e o sistema de recomendação.

Pasta “Dados” temos os arquivos utilizados para o estudo.

Pasta “Documentos” temos o cronograma de entrega do projeto e as versões de entrega deste documento.

Temos também o arquivo README.md com informações relevantes do projeto.



10.REFERENCIAL TEÓRICO