



UNIVERSIDADE PRESBITERIANA MACKENZIE
FACULDADE DE COMPUTAÇÃO E INFORMÁTICA
TECNOLOGIA EM CIÊNCIAS DE DADOS

PROJETO APLICADO III

DocuFlix: Um sistema de recomendação de filmes históricos e documentários

PROFESSORA: CAROLINA TOLEDO FERRAZ

GRUPO:

ALINE CORRÊA – 10414773 – 10414773@mackenzista.com.br

ISAQUE PIMENTEL – 10415608 – 10415608@mackenzista.com.br

MAIKI SOARES – 10415481 – 10415481@mackenzista.com.br

VANESSA CORDEIRO – 10415118 – 10415118@mackenzista.com.br

São Paulo
2024

RESUMO

Redigir ao final do trabalho.

SUMÁRIO

1.	Introdução.....	4
1.1.	Contexto do trabalho.....	4
1.2.	Motivação e justificativas.....	5
1.3.	Objetivos.....	5
1.4.	Base de dados	6
1.5.	Visão geral do pipeline.....	7
2.	Metodologia	8
3.	Resultados.....	8
4.	Conclusão e trabalhos futuros	8
5.	Referência bibliográfica.....	9

1. INTRODUÇÃO

O cinema tem um papel fundamental na sociedade, não apenas como uma forma de entretenimento, mas também como um meio de transmitir conhecimento de maneira acessível e interessante. Documentários e filmes históricos, em particular, proporcionam uma visão única sobre eventos do passado, permitindo que os espectadores absorvam informações valiosas enquanto são imersos em uma narrativa envolvente.

O objetivo deste projeto é *desenvolver um sistema de recomendação que auxilie os usuários a encontrarem filmes e documentários de alta relevância cultural e educativa*, ampliando o acesso ao aprendizado histórico por meio da sétima arte.

1.1. CONTEXTO DO TRABALHO

Somos um grupo de alunos de Ciências de Dados desenvolvendo um projeto de Sistema de Recomendação com o objetivo de melhorar as técnicas de aprendizagem de disciplinas escolares além do âmbito da sala de aula.

Com esse objetivo, propomos o **DocuFlix**, um sistema de recomendação de filmes e documentários de qualidade e relevância histórica, para estender o aprendizado da História além da aula de aula. Após o desenvolvimento do nosso produto, o apresentaremos para avaliação da disciplina de Projeto Aplicado III da Universidade Mackenzie.

1.2. MOTIVAÇÃO E JUSTIFICATIVAS

A escolha do tema é impulsionada pelo crescente interesse em métodos de ensino alternativos, que possam complementar os modelos tradicionais de educação. Filmes e documentários têm a capacidade de despertar interesse em diferentes temas, como cultura, história e ciência, através de uma abordagem lúdica e visual. Além disso, o avanço da inteligência artificial permite que sistemas de recomendação sejam utilizados para personalizar essas experiências, sugerindo conteúdos que se alinhem com as preferências dos usuários e suas necessidades de aprendizado.

A motivação pessoal e profissional para este projeto vem da necessidade de inovar na forma como o conhecimento é transmitido. Como parte do Projeto Aplicado III do curso de Ciência de Dados, este projeto permitirá a aplicação prática de técnicas de recomendação, explorando como a tecnologia pode enriquecer a experiência de aprendizagem fora do ambiente tradicional de sala de aula.

1.3. OBJETIVOS

Objetivo Geral

Desenvolver um sistema de recomendação de filmes históricos e documentários educacionais, utilizando dados de interações de usuários, com o propósito de aumentar o interesse pela história e democratizar o acesso a conteúdos audiovisuais educativos.

Objetivos Específicos

- Coletar e processar dados sobre filmes e documentários, utilizando a base de dados *MovieLens*, para criar um modelo de recomendação personalizado.
- Implementar técnicas de *filtragem colaborativa*, visando recomendar conteúdo baseado nas avaliações e preferências dos usuários.
- Propor uma solução para o problema de *Cold Start*, sugerindo conteúdos relevantes para novos usuários com poucas interações.
- Avaliar a *precisão* e a *eficácia* do sistema de recomendação, ajustando os parâmetros conforme necessário para melhorar a qualidade das sugestões.

1.4. BASE DE DADOS

A base de dados utilizada neste projeto será a **MovieLens 1M**, um conjunto de dados amplamente utilizado para pesquisas em sistemas de recomendação. A base foi criada pelo [GroupLens](#) e inclui aproximadamente 1.000.000 avaliações de filmes, fornecendo uma rica fonte de dados para o treinamento de sistema de recomendação.

Fonte dos dados

Os dados estão disponíveis publicamente em [MovieLens 1M Dataset](#).

Forma e período de coleta

Os dados foram coletados por meio das interações dos usuários da plataforma [MovieLens](#) durante o período de 2003 e 2004, onde os usuários avaliaram filmes que assistiram em uma escala de 1 a 5 estrelas.

Estrutura da base de dados

A **MovieLens 1M** inclui informações sobre:

- Usuários: Dados sobre mais de 6.000 usuários.
- Filmes: Informações sobre aproximadamente 4.000 filmes, incluindo título, ano de lançamento e gêneros.
- Avaliações: Cerca de 1.000.000 de avaliações, permitindo uma análise rica das preferências dos usuários.

Limitações da base de dados

- Diversidade de gênero: A diversidade de gêneros é limitada a filmes listados na base, o que pode obviamente afetar a variedade de recomendações.
- Antiguidade dos dados: A coleta de dados é antiga, e muitos filmes recentes não estão incluídos, o que pode reduzir a relevância de algumas recomendações.
- Cold Start: A base não tem informações ricas sobre novos usuários ou filmes, o que pode dificultar a solução do problema de Cold Start.

1.5. VISÃO GERAL DO PIPELINE

Este projeto busca aplicar técnicas de Ciência de Dados para desenvolver um sistema de recomendação que enriqueça a experiência de aprendizado sobre História, ampliando o acesso a conteúdos audiovisuais de qualidade. Segue abaixo a visão geral do pipeline a ser adotado:

Coleta de dados

A base MovieLens 1M será carregada e processada para obter informações sobre filmes, avaliações e interações dos usuários.

Pré-processamento

Primeiramente, será realizado o tratamento de dados faltantes. Depois, será aplicada uma normalização das avaliações. Finalmente, será feita a conversão da matriz usuário-item para um formato adequado para o treinamento do modelo.

Construção do modelo de recomendação:

Iremos realizar a implementação da técnica de Filtragem Colaborativa, utilizando a Fatoração de Matriz. Depois realizaremos o treinamento do modelo com as avaliações dos usuários em máquinas locais.

Solução do problema Cold Start

Para novos usuários, será feita a recomendação de filmes e documentários populares ou de alta avaliação.

Avaliação e ajustes:

Será necessário testar a acurácia do sistema utilizando métricas como **RMSE (Root Mean Squared Error)**. Por fim, iremos ajustar hiperparâmetros e otimizar o modelo conforme necessário.

2. METODOLOGIA

Redigir a partir da Aula 2.

3. RESULTADOS

Redigir a partir da aula 2.

4. CONCLUSÃO E TRABALHOS FUTUROS

Redigir a partir da aula 2.

5. REFERÊNCIA BIBLIOGRÁFICA

F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 4, Article 19 (December 2015), 19 pages. DOI=<http://dx.doi.org/10.1145/2827872>.