



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

## **PROJETO APLICADO III**

**HistFlix: Um sistema de recomendação personalizado de filmes  
históricos e documentários**

**PROFESSORA:** CAROLINA TOLEDO FERRAZ

**GRUPO:**

BRUNO BALTUILHE – 10424822 – 10424822@mackenzista.com.br

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

KELLY GRAZIELY PENA – 110416108 – 10416108@mackenzista.com.br

São Paulo  
2025

## **RESUMO**

Redigir ao final do trabalho.

# SUMÁRIO

1.	Introdução.....	5
1.1.	Contexto do trabalho.....	5
1.2.	Motivação e justificativas.....	6
1.3.	Objetivos.....	7
2.	Referencial Teórico.....	8
2.1.	Análise Exploratória e Preparação dos Dados .....	8
2.2.	Sistemas de Recomendação e Modelo Híbrido.....	8
2.3.	Personalização Contextual com NLP e Análise Emocional .....	9
2.4.	Aplicação no HistFlix.....	10
2.5.	Avaliação do Sistema .....	10
3.	METODOLOGIA.....	11
3.1.	Novo Treinamento e Reavaliação do Modelo.....	11
3.2.	Técnicas Utilizadas e a Utilizar .....	11
3.3.	Etapas Realizadas e Dificuldades Superadas .....	12
3.4.	Contribuição para o campo de pesquisa “Sistemas de Recomendação” .....	13
4.	Referência bibliográfica.....	14
5.	Anexo .....	16
5.1.	Definição das Bibliotecas Python .....	16



## 1. INTRODUÇÃO

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

O consumo de filmes, séries e documentários tem se tornado cada vez mais personalizado, impulsionado pelo avanço da tecnologia e pelos sistemas de recomendação. Essa personalização melhora a experiência do usuário e amplia o acesso a conteúdos relevantes que poderiam passar despercebidos.

O objetivo deste projeto <sup>1</sup>é **desenvolver um sistema de recomendação que auxilie os usuários a encontrarem filmes e documentários de alta relevância cultural e educativa, considerando não apenas preferências passadas, mas também fatores contextuais como estado emocional**, ampliando o acesso ao aprendizado histórico por meio da sétima arte e proporcionando recomendações mais adequadas e personalizadas.

### 1.1. CONTEXTO DO TRABALHO

Somos um grupo de alunos de Ciências de Dados desenvolvendo um projeto de Sistema de Recomendação para melhorar as técnicas aprendizagem de disciplinas escolares além da sala de aula. Propomos o **HistFlix**, um sistema de recomendação de filmes e documentários de qualidade e relevância histórica e educacional, para estender o aprendizado da História além da sala de aula. Após o desenvolvimento do nosso produto, o apresentaremos para avaliação da disciplina de Projeto Aplicado III da Universidade Mackenzie.

---

<sup>1</sup> Segue a URL do projeto no Github: <https://github.com/isaque-pimentel/projeto-aplicado-3>

## 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, séries e documentários despertam 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 personalizem essas experiências, sugerindo conteúdos alinhados às preferências dos usuários e suas necessidades de aprendizado.

Este projeto também busca aprimorar os sistemas de recomendação existentes, tornando-os mais dinâmicos e sensíveis ao contexto do usuário. Ao incorporar inteligência artificial para interpretar emoções expressadas por meio de uma interação textual, o sistema poderá fornecer sugestões mais alinhadas às necessidades específicas de cada momento. Isso contribui para um consumo mais significativo e envolvente de conteúdos audiovisuais, reduzindo a frustração causada por recomendações irrelevantes.

A *HistFlix* busca suprir a necessidade de um sistema especializado que forneça recomendações precisas e relevantes para estudantes, pesquisadores e entusiastas da história. Além disso, o projeto alinha-se aos Objetivos de Desenvolvimento Sustentável (ODS) da ONU, promovendo educação de qualidade ao facilitar o acesso a conteúdos educativos.

## 1.3. OBJETIVOS

### Objetivo Geral

Desenvolver um sistema de recomendação de filmes, séries e documentários que utilize um modelo híbrido de recomendação e inteligência artificial para interpretar as emoções do usuário e sugerir conteúdos alinhados ao seu estado emocional e preferências específicas, 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, séries e documentários, utilizando a base de dados **MovieLens**, para criar um modelo de recomendação personalizado.
- Implementar um modelo híbrido de recomendação, combinando filtragem colaborativa (que analisa o comportamento e as avaliações de outros usuários com perfis semelhantes) e filtragem baseada em conteúdo (que considera características específicas das obras audiovisuais, como gênero, duração, elenco e temática).
- Desenvolver uma interface interativa na qual os usuários possam expressar suas emoções e preferências momentâneas.
- Integrar técnicas de Processamento de Linguagem Natural (PLN) para interpretar sentimentos e preferências expressas textualmente.

## 2. REFERENCIAL TEÓRICO

Este referencial teórico sustenta as decisões técnicas e metodológicas do *HistFlix*, nosso sistema inteligente de recomendação de conteúdos audiovisuais históricos, com sensibilidade ao estado emocional dos usuários. A proposta combina modelos híbridos de recomendação com técnicas de Processamento de Linguagem Natural (NLP), promovendo uma experiência personalizada, educativa e contextualizada.

### 2.1. ANÁLISE EXPLORATÓRIA E PREPARAÇÃO DOS DADOS

Para o desenvolvimento do sistema, utilizou-se a base **MovieLens**, amplamente adotada em pesquisas sobre sistemas de recomendação. Os dados originais, armazenados em arquivos *.dat*, foram transformados em *DataFrames* com a biblioteca *pandas*, seguindo as etapas:

- Normalização e limpeza dos dados, incluindo remoção de inconsistências e codificação de variáveis categóricas;
- Extração de atributos adicionais (ex.: ano de lançamento dos filmes);
- Armazenamento estruturado em um banco de dados relacional SQLite com tabelas para usuários, filmes e avaliações.

Em seguida, foi conduzida uma **Análise Exploratória de Dados (EDA)** com foco na distribuição das notas, perfis de usuários, frequência dos gêneros e evolução temporal das avaliações. Essa preparação foi essencial para garantir a integridade dos dados e orientar a escolha das técnicas de recomendação.

### 2.2. SISTEMAS DE RECOMENDAÇÃO E MODELO HÍBRIDO

Sistemas de recomendação são ferramentas computacionais que sugerem itens com base em preferências passadas, características de conteúdo ou comportamentos de usuários similares (Ricci et al., 2015). As abordagens mais comuns são a Filtragem Colaborativa (FC) e a Filtragem Baseada em Conteúdo (FBC).

A FC, especialmente em sua forma baseada em fatoração de matrizes, como o Singular Value Decomposition (SVD), é eficaz em bases densas como a MovieLens,



ao identificar padrões entre usuários e itens (Koren et al., 2009; Harper & Konstan, 2015). No entanto, sofre com o problema do *cold start* para novos usuários ou itens pouco avaliados.

A FBC, por sua vez, considera atributos dos itens (como gêneros ou palavras-chave), sendo útil em situações com pouca interação histórica. Técnicas como TF-IDF e similaridade do cosseno são amplamente empregadas para analisar descrições e metadados (Lops et al., 2011; Aggarwal, 2016).

A combinação das abordagens é realizada por meio de **score-level fusion** (Burke, 2002), segundo a equação:

$$\text{score final} = \alpha \cdot \text{score}_{\text{FC}} + (1 - \alpha) \cdot \alpha \cdot \text{score}_{\text{FBC}}$$

em que  $\alpha$  representa o peso ajustável entre as duas fontes de recomendação. Estudos recentes, como Campos et al. (2021), apontam que modelos híbridos têm grande potencial em contextos educacionais e culturais, especialmente quando adaptados à diversidade de perfis e interesses.

### 2.3. PERSONALIZAÇÃO CONTEXTUAL COM NLP E ANÁLISE EMOCIONAL

A integração de técnicas de NLP permite ao sistema interpretar entradas textuais dos usuários, extraíndo emoções e tópicos de interesse. Ferramentas como **TextBlob** e **VADER** são utilizadas para análise de sentimentos em linguagem informal, contribuindo para a construção de perfis mais sensíveis ao contexto emocional.

Trabalhos como **RecEmo** (Zhao et al., 2021) e *Emotional Recommender System* (Wang et al., 2023) demonstram como o uso de emoções pode aprimorar a relevância das recomendações, adaptando-as ao humor ou intenção do usuário. O levantamento feito por Chen, Li e Wang (2020) reforça essa tendência ao revisar abordagens de recomendação sensíveis à emoção e propor direções futuras, consolidando o valor da personalização emocional em sistemas de recomendação modernos.

Além disso, estudos como o DeepMovie (Kumar & Sharma, 2020) mostram o potencial da combinação entre NLP e análise de sentimentos em sistemas voltados

para recomendação de filmes, utilizando redes neurais profundas para melhorar a precisão e a sensibilidade emocional das sugestões.

No campo educacional, o EduFlix (Silva et al., 2022) integrou metadados e NLP para apoiar o aprendizado por meio de vídeos, inspirando diretamente a proposta do *HistFlix*.

## 2.4. APLICAÇÃO NO HISTFLIX

O *HistFlix* adota um modelo híbrido sensível ao contexto, integrando FC (via SVD), FBC (via TF-IDF e similaridade do cosseno) e uma camada de personalização emocional com NLP. O sistema processa frases livres digitadas pelos usuários para identificar emoções predominantes e temas históricos. Com base nisso, as recomendações são geradas e reordenadas conforme a compatibilidade emocional dos conteúdos.

Essa abordagem visa aumentar a precisão técnica das sugestões, e promover um engajamento mais significativo com conteúdos históricos, alinhando-se à proposta educacional do projeto.

## 2.5. AVALIAÇÃO DO SISTEMA

A eficácia do *HistFlix* será avaliada por meio de métricas clássicas da literatura, incluindo:

- **Root Mean Squared Error (RMSE):** para medir a precisão das predições numéricas;
- **Precision@K e Recall@K:** para avaliar a relevância dos itens recomendados entre os top-K resultados.

Essas métricas permitem validar a qualidade técnica do sistema e sua utilidade prática na interface do usuário.

### 3. METODOLOGIA

A prova de conceito mostrou que o modelo híbrido, combinando filtragem colaborativa, baseada em conteúdo, funcionava bem para usuários com histórico amplo. No entanto, observou-se baixa performance para novos usuários e baixa cobertura do catálogo. As métricas de desempenho apontaram a necessidade de melhorias no pipeline.

Para melhorar o desempenho do sistema, será necessário organizar melhor a base de dados, garantindo que o acesso às informações seja mais rápido e eficiente. Em seguida, é importante enriquecer essa base com novos dados sobre os filmes, o que vai permitir ao sistema compreender melhor os gostos e padrões dos usuários.

Outro passo fundamental seria ajustar o modo como o sistema escolhe o que recomendar, buscando um equilíbrio mais inteligente entre as sugestões baseadas no histórico de cada usuário e aquelas geradas pelas características dos próprios filmes, e somando a filtragem mista que estamos criando, tornam as recomendações mais alinhadas com o que ele realmente procura assistir naquele momento.

#### 3.1. NOVO TREINAMENTO E REAVALIAÇÃO DO MODELO

Depois dos ajustes no sistema, notamos que é preciso fazer um novo treinamento para ver se as mudanças realmente melhoraram as recomendações. Com a base de dados mais limpa e organizada, o sistema deve entender melhor o que mostrar para cada tipo de usuário. As emoções informadas pelos usuários também serão interpretadas de forma mais precisa, o que ajudará a personalizar ainda mais as sugestões.

Usamos alguns testes para comparar os resultados com a versão anterior. O sistema conseguiu recomendar mais filmes diferentes, aumentando a cobertura e recuperou melhor os títulos relevantes melhorando o recall.

#### 3.2. TÉCNICAS UTILIZADAS

**Filtragem Colaborativa:** o sistema aprende com as preferências de outros usuários com gostos parecidos para sugerir filmes.

- **Filtragem Baseada em Conteúdo:** analisa informações dos próprios filmes, como gênero e duração, para fazer recomendações que combinem com os interesses do usuário.

Ainda não implementamos a última técnica de Processamento de Linguagem Natural.

### 3.3. ETAPAS REALIZADAS E DIFICULDADES SUPERADAS

Durante o desenvolvimento do HistFlix, enfrentamos desafios práticos que exigiram ajustes constantes ao longo do processo. A estruturação da base MovieLens 1M revelou limitações, como a falta de atualizações e uma diversidade emocional reduzida.

No pré-processamento, padronizamos os dados textuais e lidamos com o desequilíbrio entre os modelos de recomendação, ajustando a ponderação entre filtragem colaborativa e baseada em conteúdo para obter resultados mais equilibrados. No entanto, o problema de cold start, causado pela ausência de histórico em perfis novos, ainda persiste e está sendo analisado para futuras melhorias. A abordagem atual considera o uso de dados derivados e características de perfil como alternativas iniciais para mitigar esse impacto.

Além disso, começamos a implementar um analisador de sentimentos baseado no TextBlob, mas ele ainda está em fase inicial de calibração. A interpretação emocional dos dados segue sendo ajustada para garantir que as recomendações reflitam de forma mais precisa o estado do usuário e proporcionem uma experiência mais personalizada.

Embora esses desafios ainda não tenham sido completamente superados, o progresso contínuo e os testes iterativos vêm fortalecendo o pipeline do projeto, tornando-o mais robusto e adaptável. As próximas etapas incluem refinamentos na modelagem preditiva e aprimoramento dos mecanismos de recomendação para enfrentar as dificuldades remanescentes.

### 3.4. CONTRIBUIÇÃO PARA O CAMPO DE PESQUISA “SISTEMAS DE RECOMENDAÇÃO”

O **HistFlix** apresenta uma proposta inovadora ao unir sistemas de recomendação com análise de sentimentos, oferecendo uma experiência mais próxima da realidade de quem consome conteúdos educacionais e históricos. Ao considerar não só as preferências anteriores, mas também o estado emocional do usuário, o projeto mostra um caminho promissor para tornar as recomendações mais humanas e eficazes.

Com essa abordagem, o **HistFlix** amplia as possibilidades dentro da área de sistemas de recomendação, trazendo uma nova perspectiva sobre personalização. Ao integrar emoções ao processo de sugestão de conteúdo, mostramos que é possível ir além das preferências técnicas e criar sistemas que se adaptam de forma mais sensível às pessoas. Isso abre espaço para pesquisas futuras, inspira melhorias em soluções já existentes e contribui com uma proposta concreta que pode ser adaptada em outros projetos e áreas de aplicação.

#### 4. 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>.

Chongchun Aggarwal. 2016. *Recommender Systems: The Textbook*. Springer. DOI=<https://doi.org/10.1007/978-3-319-29659-3>.

Robin Burke. 2002. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12, 331–370. DOI=<https://doi.org/10.1023/A:1021240730564>.

Pedro Campos, Fernando Díez, and Iván Cantador. 2021. Personalized and Adaptive Learning Systems Using Hybrid Recommender Models. *International Journal of Artificial Intelligence in Education*, 31, 845-867. DOI=<https://doi.org/10.1007/s40593-021-00258-9>.

Yong Chen, Jie Li, and Xiaojun Wang. 2020. Emotion-aware Recommendation: A Survey and Future Directions. *ACM Transactions on Information Systems (TOIS)* 38, 4, Article 34 (July 2020), 34 pages. DOI=<https://doi.org/10.1145/3394621>.

Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer*, 42, 8, 30-37. DOI=<https://doi.org/10.1109/MC.2009.263>.

Ravi Kumar and Anuj Sharma. 2020. DeepMovie: A Deep Learning-based Movie Recommender System Using NLP and Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 11, 8, 317-325. DOI=<https://doi.org/10.14569/IJACSA.2020.0110839>.

Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2011. Content-Based Recommender Systems: State of the Art and Trends. In *Recommender*

Systems Handbook, Springer, 73-105. DOI=[https://doi.org/10.1007/978-0-387-85820-3\\_3](https://doi.org/10.1007/978-0-387-85820-3_3).

Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender Systems Handbook (2nd ed.). Springer. DOI=<https://doi.org/10.1007/978-1-4899-7637-6>.

Marcelo Silva, Pedro Santos, and Rafael Almeida. 2022. EduFlix: A Recommender System for Educational Videos Based on Metadata and NLP Techniques. Journal of Educational Technology & Society, 25, 3, 117-132.

Lei Wang, Xiaojun Chen, and Yuhao Zhao. 2023. Emotional Recommender System: An Approach to Personalized Content Suggestions Based on User Mood. Journal of Artificial Intelligence Research, 76, 321-345. DOI=<https://doi.org/10.1613/jair.2023.124>.

Yunhao Zhao, Jian Huang, and Wei Lin. 2021. RecEmo: Emotion-Driven Content Recommendation Using Sentiment Analysis. IEEE Transactions on Affective Computing, 12, 2, 387-398. DOI=<https://doi.org/10.1109/TAFFC.2021.3072674>.

## 5. ANEXO

### 5.1. DEFINIÇÃO DAS BIBLIOTECAS PYTHON

Bibliotecas são fundamentais para a construção de qualquer software robusto e eficiente. No desenvolvimento do nosso sistema de recomendação, diversas bibliotecas Python foram empregadas para análise de dados, visualização, criação de recomendações e processamento de linguagem natural:

#### Manipulação e Análise de Dados

- **pandas**: manipula dados em formato de tabelas (chamados de DataFrames), organizando os dados da base MovieLens, como usuários, filmes e avaliações.
- **numpy**: realiza cálculos matemáticos eficientes com vetores e matrizes, auxiliando em operações numéricas e estatísticas, como normalização.

#### Visualização de Dados

- **matplotlib**: cria gráficos básicos, como histogramas e curvas de erro, úteis para exibição de desempenho do modelo.
- **seaborn**: complementa o matplotlib com gráficos mais intuitivos, como heatmaps e boxplots, facilitando a análise exploratória.

#### Sistema de Recomendação

- **sklearn.metrics.pairwise.cosine\_similarity**: mede a similaridade entre vetores, essencial para calcular proximidade entre usuários ou filmes.
- **sklearn.feature\_extraction.text.TfidfVectorizer**: converte descrições de filmes em vetores numéricos (TF-IDF), permitindo recomendação baseada em conteúdo.
- **surprise.SVD**: implementa filtragem colaborativa por meio de fatoração de matriz, prevendo avaliações de usuários.
- **surprise.Dataset** e **surprise.Reader**: estruturam dados no formato adequado para alimentar modelos de recomendação.



## Processamento de Linguagem Natural

- **TextBlob**: analisa sentimentos e interpreta emoções dos textos inseridos, como "triste" ou "animado", personalizando recomendações.
- **nltk**: realiza tokenização e lematização, além de remover stopwords, facilitando o pré-processamento de textos antes da análise.