



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

Este trabalho apresenta o desenvolvimento do **HistFlix**, um sistema inteligente de recomendação de filmes históricos e documentários com sensibilidade ao estado emocional dos usuários. A proposta combina técnicas de sistemas de recomendação híbridos, análise de sentimentos por meio de Processamento de Linguagem Natural (PLN) e uma interface orientada à personalização no contexto educacional. Para isso, utilizou-se a base de dados **MovieLens 1M**, que foi enriquecida com dados emocionais inferidos de interações simuladas. O sistema emprega uma abordagem de fusão entre filtragem colaborativa e baseada em conteúdo, ajustada com base no perfil e no humor do usuário. A análise crítica dos desafios enfrentados, como o problema de *cold start* e limitações na análise de sentimentos, orienta as propostas de trabalhos futuros. O **HistFlix** representa uma contribuição original ao unir aspectos afetivos e pedagógicos em sistemas de recomendação, com potencial aplicação em ambientes de aprendizagem personalizados.

# SUMÁRIO

1.	Introdução .....	5
1.1.	Contexto do trabalho .....	5
1.2.	Motivação e justificativas.....	5
1.3.	Objetivos.....	6
2.	Referencial Teórico .....	7
2.1.	Sistemas de Recomendação Híbridos .....	7
2.2.	Personalização Contextual com NLP e Análise Emocional .....	8
3.	Metodologia.....	9
3.1.	Coleta e Processamento dos Dados.....	9
3.2.	Modelagem do Sistema de Recomendação.....	10
3.3.	Análise de Sentimentos com PLN .....	10
3.4.	Desenvolvimento da Interface.....	11
4.	Resultados .....	12
4.1.	Etapas Realizadas e Dificuldades Superadas .....	12
4.2.	Síntese dos Resultados e Perspectivas .....	12
5.	Conclusão e Trabalhos Futuros .....	14
6.	Referência bibliográfica.....	15
7.	Anexo .....	17
7.1.	Definição das Bibliotecas Python .....	17



## 1. INTRODUÇÃO

O cinema desempenha um papel fundamental na sociedade, não apenas como forma de entretenimento, mas também como meio de transmitir conhecimento de maneira acessível e envolvente. Filmes, séries e documentários, especialmente os de cunho histórico, oferecem uma perspectiva única sobre eventos do passado, promovendo aprendizado de forma lúdica e visual.

O consumo de conteúdos audiovisuais tem se tornado cada vez mais personalizado, impulsionado pelo avanço dos sistemas de recomendação. Essa personalização melhora a experiência do usuário e amplia o acesso a conteúdos relevantes que, de outra forma, poderiam passar despercebidos. Tais sistemas são particularmente promissores quando aplicados ao contexto educacional, contribuindo para o engajamento e a aprendizagem fora do ambiente escolar tradicional.

Este trabalho propõe o desenvolvimento do **HistFlix**<sup>1</sup>, um sistema de recomendação híbrido que sugere filmes e documentários de alta relevância histórica e educativa. O sistema combina técnicas de filtragem colaborativa e baseada em conteúdo, além de incorporar dados contextuais como o estado emocional do usuário, com o objetivo de tornar as recomendações mais personalizadas e eficazes.

### 1.1. CONTEXTO DO TRABALHO

O projeto foi desenvolvido no âmbito da disciplina de Projeto Aplicado III do curso de Ciência de Dados da Universidade Presbiteriana Mackenzie. O objetivo central é utilizar sistemas inteligentes para promover o ensino de História por meio de recursos audiovisuais de qualidade. O **HistFlix** busca atender estudantes, professores e entusiastas da área que desejam aprofundar seus conhecimentos fora do ambiente formal de ensino.

### 1.2. MOTIVAÇÃO E JUSTIFICATIVAS

O interesse crescente em métodos de ensino alternativos tem impulsionado a adoção de recursos como filmes e documentários para complementar os modelos

---

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

educacionais tradicionais. Esses conteúdos despertam o interesse por temas complexos, como eventos históricos, por meio de narrativas visuais mais acessíveis.

Além disso, os avanços em inteligência artificial possibilitam a criação de sistemas de recomendação cada vez mais precisos e sensíveis ao contexto do usuário. Este projeto busca aprimorar tais sistemas por meio da integração de técnicas de Processamento de Linguagem Natural (PLN) para interpretar emoções expressas textualmente, o que permite fornecer recomendações mais alinhadas às necessidades e ao estado emocional do usuário.

O **HistFlix** visa preencher a lacuna existente na oferta de sistemas de recomendação especializados em conteúdos educativos com foco em História, promovendo o acesso democrático à educação de qualidade. O projeto também está alinhado com os Objetivos de Desenvolvimento Sustentável (ODS) da ONU, especialmente o ODS 4, que trata da educação de qualidade.

### 1.3. OBJETIVOS

#### Objetivo Geral

Desenvolver um sistema de recomendação híbrido que utilize inteligência artificial para interpretar emoções e preferências dos usuários, com foco em conteúdos audiovisuais de caráter histórico e educativo, promovendo o engajamento e o acesso ao conhecimento.

#### Objetivos Específicos

- Coletar e processar dados sobre filmes, séries e documentários por meio da base **MovieLens**;
- Implementar um modelo híbrido de recomendação que combine filtragem colaborativa e filtragem baseada em conteúdo;
- Desenvolver uma interface interativa para coleta de emoções e preferências momentâneas dos usuários;
- Integrar técnicas de Processamento de Linguagem Natural (PLN) para interpretar sentimentos expressos em texto;

## 2. REFERENCIAL TEÓRICO

Este referencial teórico sustenta as decisões técnicas e metodológicas adotadas no desenvolvimento do **HistFlix**, um sistema inteligente de recomendação de conteúdos audiovisuais históricos, sensível ao estado emocional do usuário. A proposta combina técnicas híbridas de recomendação com métodos de Processamento de Linguagem Natural (PLN), promovendo uma experiência personalizada, educativa e contextualizada.

### 2.1. SISTEMAS DE RECOMENDAÇÃO HÍBRIDOS

Sistemas de recomendação são amplamente utilizados para auxiliar usuários na descoberta de itens relevantes, com base em dados históricos, características de conteúdo ou similaridade entre perfis. As abordagens mais comuns, e que serão usadas no desenvolvimento do **HistFlix**, incluem a *Filtragem Colaborativa* (FC) e a *Filtragem Baseada em Conteúdo* (FBC) (Ricci et al., 2022).

A FC, especialmente em sua forma baseada em fatoração de matrizes, como o *Singular Value Decomposition* (SVD), tem se mostrado eficaz em conjuntos de dados densos, como o MovieLens (Harper & Konstan, 2015), ao identificar padrões latentes entre usuários e itens (Koren et al., 2009). Contudo, esse método enfrenta dificuldades no cenário do *cold start*, quando há pouca informação sobre novos usuários ou itens.

Por outro lado, a FBC analisa atributos explícitos dos itens, como gêneros, diretor ou palavras-chave. Métodos como *Term Frequency-Inverse Document Frequency* (TF-IDF) e similaridade do cosseno são comumente utilizados para esse fim (Lops et al., 2011; Aggarwal, 2016).

Para mitigar as limitações de cada abordagem isoladamente, são empregados modelos híbridos que combinam os scores individuais. A combinação pode ser realizada por meio da seguinte 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 destacam o potencial dos modelos híbridos em contextos educacionais, culturais e sensíveis à diversidade de perfis (Campos et al., 2021).

## 2.2. PERSONALIZAÇÃO CONTEXTUAL COM NLP E ANÁLISE EMOCIONAL

Com a evolução das técnicas de PLN, torna-se possível incorporar variáveis contextuais e emocionais aos sistemas de recomendação. Ferramentas como **VADER** e **TextBlob** têm sido utilizadas para extrair sentimentos de entradas textuais em linguagem natural, especialmente em ambientes digitais e redes sociais (Hutto et al., 2014).

Modelos mais recentes, como o **RecEmo** (Zhao et al., 2021) e *Emotional Recommender System* (Wang et al., 2023) mostram que a inclusão de sentimentos pode aumentar significativamente a relevância das recomendações, ao alinhar sugestões com o estado emocional momentâneo do usuário.

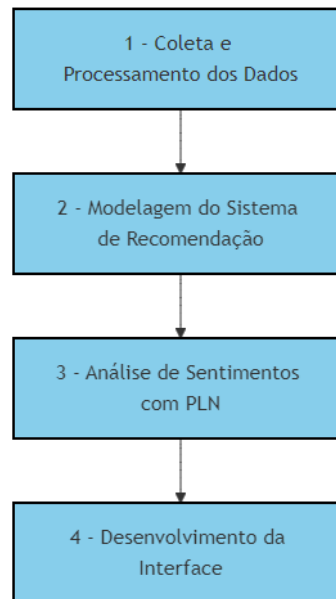
Revisões sistemáticas, como a de Chen, Li e Wang (2020), reforçam a importância da personalização emocional, destacando seu impacto positivo na satisfação e no engajamento dos usuários. Além disso, trabalhos como o **DeepMovie** (Kumar & Sharma, 2020) utilizam redes neurais profundas para integrar dados de sentimento e características dos filmes, aumentando a precisão das recomendações.

No campo educacional, o projeto **EduFlix** (Silva et al., 2022) demonstrou o valor da integração entre metadados audiovisuais e técnicas de PLN para promover o aprendizado via vídeos educacionais, inspirando diretamente o conceito do **HistFlix**.



### 3. METODOLOGIA

Aqui são descritas as etapas metodológicas adotadas no desenvolvimento do **HistFlix**, desde a seleção da base de dados até a implementação do modelo de recomendação e da interface interativa. A abordagem segue um modelo incremental, com ciclos de teste e validação conforme o fluxograma apresentado abaixo.



#### 3.1. COLETA E PROCESSAMENTO 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.

### 3.2. MODELAGEM DO SISTEMA DE RECOMENDAÇÃO

O **HistFlix** adotará um *modelo híbrido* sensível ao contexto, combinando:

- **Filtragem Colaborativa** baseada na fatoração de matrizes via *Singular Value Decomposition* (SVD);
- **Filtragem Baseada em Conteúdo**, com vetores TF-IDF dos nomes e gêneros dos filmes;
- **Combinação por score-level fusion**, ajustando  $\alpha$  empiricamente para maximizar a precisão top-K.

A eficácia do **HistFlix** será avaliada por meio de métricas clássicas da literatura, incluindo: **Root Mean Squared Error (RMSE)** para medir a precisão das previsões numéricas, **Precision@K** e **Recall@K** para avaliar a relevância dos itens recomendados entre os top-K resultados.

Essas métricas permitem ajustar o modelo através da validação cruzada com *handout* de 20% dos dados de avaliação, garantindo a qualidade técnica do sistema.

### 3.3. ANÁLISE DE SENTIMENTOS COM PLN

O sistema incluirá um módulo de interpretação emocional com NLP, onde o sistema processa frases livres digitadas pelos usuários para identificar emoções predominantes e temas históricos (como “Estou animado para aprender sobre guerras antigas”). Serão utilizados os seguintes passos:

- Análise de sentimento utilizando TextBlob;
- Extração de palavras-chave por TF-IDF;
- Mapeamento de emoções para categorias como: *curioso*, *reflexivo*, *nostálgico* etc.

O resultado dessa análise será usado para ajustar o score final de recomendação, atribuindo maior peso a conteúdos historicamente relevantes compatíveis com o estado emocional identificado.

### 3.4. DESENVOLVIMENTO DA INTERFACE

A interface será desenvolvida em ambiente web, utilizando **Python com Flask**, permitindo:

- Campo de entrada textual para expressão do estado emocional;
- Exibição de recomendações personalizadas em tempo real.

## 4. RESULTADOS

Aqui são apresentados os principais achados decorrentes da implementação e testes do sistema **HistFlix**, incluindo as melhorias observadas após ajustes no modelo híbrido, os desafios enfrentados ao longo do desenvolvimento e os avanços obtidos. Os resultados são analisados de forma qualitativa, considerando tanto aspectos funcionais quanto técnicos.

### 4.1. ETAPAS REALIZADAS E DIFICULDADES SUPERADAS

Durante o desenvolvimento do **HistFlix**, foram enfrentados diversos desafios técnicos e metodológicos. Abaixo, destacam-se os principais pontos enfrentados:

- **Base de dados:** a estruturação da **MovieLens 1M** revelou limitações quanto à atualização temporal e à diversidade emocional. Foi necessário realizar pré-processamento extensivo, incluindo enriquecimento de metadados.
- **Desbalanceamento entre técnicas de recomendação:** a combinação entre filtragem colaborativa e baseada em conteúdo exigiu ajuste fino do parâmetro de ponderação, a fim de equilibrar a contribuição de cada técnica. Isso permitiu reduzir o viés das recomendações mais populares.
- **Análise emocional inicial:** a implementação do analisador de sentimentos, com **TextBlob**, ainda se encontra em fase de calibração. A precisão na identificação de emoções e sua correlação com recomendações é um ponto de contínua iteração.

Embora ainda haja pontos de aprimoramento, o sistema tem evoluído por meio de ciclos iterativos de desenvolvimento e testes.

### 4.2. SÍNTESE DOS RESULTADOS E PERSPECTIVAS

Os principais pontos positivos observados até o momento incluem:

- Maior capacidade de personalização das recomendações;
- Integração funcional entre análise de sentimentos e motor de recomendação;
- Melhor cobertura e recuperação de conteúdos relevantes.

Entre os pontos a serem aperfeiçoados destacam-se:

- Aprimoramento da análise emocional com modelos mais robustos de PLN;
- Enriquecimento da base de dados com atributos emocionais e educacionais mais diversificados;
- Soluções mais eficazes para o problema de *cold start*.

Os resultados parciais são promissores e indicam que o **HistFlix** possui potencial para se tornar uma ferramenta no apoio ao ensino de História.

## 5. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou o desenvolvimento do **HistFlix**, um sistema inteligente de recomendação de conteúdos audiovisuais históricos, com sensibilidade ao estado emocional dos usuários. A proposta integra técnicas de sistemas de recomendação híbridos, análise de sentimentos com Processamento de Linguagem Natural (PLN) e uma interface amigável voltada ao uso educacional.

Os principais resultados obtidos indicam que a inclusão da variável emocional nas recomendações ampliou a personalização e melhorou o alinhamento entre o conteúdo sugerido e o contexto subjetivo do usuário. A reavaliação do modelo após ajustes no pipeline evidenciou um avanço qualitativo na relevância das recomendações.

A contribuição central deste estudo reside na incorporação de aspectos afetivos e semânticos no processo de recomendação, ampliando o escopo dos sistemas tradicionais centrados exclusivamente em preferências históricas ou similaridade de conteúdo. O **HistFlix** propõe uma abordagem inovadora para o uso da inteligência artificial no apoio ao ensino de História, promovendo uma experiência mais sensível, contextualizada e engajadora.

No entanto, desafios persistem, como o tratamento do problema de *cold start* e a limitação da base de dados utilizada, que carece de diversidade emocional e atualizações mais recentes. A análise de sentimentos, por ora implementada com ferramentas básicas como o **TextBlob**, também requer refinamento e validação com modelos mais robustos.

Como trabalhos futuros, propõe-se:

- A adoção de modelos de linguagem mais avançados, como BERT ou RoBERTa, para análise emocional e extração semântica;
- A expansão e atualização da base de dados com conteúdos mais diversos e anotados emocionalmente;
- A realização de testes com usuários reais, visando validar empiricamente a eficácia da recomendação emocional no contexto educacional.

## 6. 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). 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>.

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). 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. 2022. *Recommender Systems Handbook* (2nd ed.). Springer. DOI=<https://doi.org/10.1007/978-1-0716-2197-4>.

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>.



## 7. ANEXO

### 7.1. 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.