

## Recomendação Personalizada de Filmes: Desenvolvimento de Software e Prova de Conceito de Feature Weighting (Utilizando Streamlit)

**Glenda Souza Fernandes Dos Santos; Isaac Oliveira Dias; João Luccas Lordelo Marques; Marcus Vinicius Lameu Lima; Paulo Henrique Pereira Araujo Piedade; Adailton de Jesus Cerqueira Junior.**

Universidade Salvador – Salvador – BA – Brasil

### INTRODUÇÃO

O desafio da sobrecarga de informação em plataformas de mídia é mitigado pelos Sistemas de Recomendação (SR). Este trabalho foca na arquitetura Baseada em Conteúdo (Content-Based), que frequentemente falha ao atribuir pesos uniformes às features dos itens.

O problema de pesquisa é como otimizar a precisão do SR Content-Based através da ponderação dinâmica e estratégica de características. O Objetivo Geral é desenvolver um sistema Content-Based aprimorado com um módulo de Ponderação de Características (Feature Weighting) e uma interface web funcional.

A metodologia é de Desenvolvimento Tecnológico e Experimental, utilizando Similaridade de Cosseno sobre vetores criados via One-Hot Encoding (dados IMDB). O projeto é dividido em Treinamento (construção da matriz de similaridade) e Inferência (aplicação dos pesos dinâmicos via Streamlit).

### MÉTODOS

**Treinamento:** O script inicial (treinamento\_modelo\_filmes.py) realiza o pré-processamento de dados do IMDB, incluindo a Codificação One-Hot para transformar características categóricas (Gênero, Diretor, Ator, etc.) em vetores numéricos. Em seguida, calcula a Matriz de Similaridade de Cosseno entre os filmes, essencial para o algoritmo.  
**Inferência e Usabilidade:** O aplicativo (app\_filmes.py), construído com Streamlit, implementa a fase de recomendação em tempo real (Usuário-para-Item). O sistema aplica a melhoria metodológica de Ponderação de Características (Feature Weighting), atribuindo pesos variáveis (e.g., Gênero pesa mais que Produtora) de forma dinâmica, conforme a seleção do usuário, para otimizar a precisão das sugestões.  
**Embásamento:** A arquitetura é fundamentada na Modelagem de Itens e na Vetorização de Características, onde a similaridade é medida pela proximidade dos vetores no espaço multidimensional.

### RESULTADOS

Esta seção apresenta e analisa o desempenho do Sistema de Recomendação Content-Based aprimorado. A coerência lógica das recomendações foi avaliada utilizando a Média de Jaccard, que compara as características definidas nos Perfis Sintéticos com as características presentes nos filmes sugeridos. O índice médio encontrado foi de 0,2513 (25,13%), o que indica que aproximadamente um quarto das features fornecidas pelo usuário aparecem, em média, nos filmes recomendados. Observou-se uma diferença clara entre os tipos de perfis avaliados: os Perfis Comuns, baseados em categorias frequentes no conjunto de dados, apresentaram os melhores índices de coerência, pois o modelo encontra diversas instâncias semelhantes no treinamento, resultando em maior similaridade. Em contraste, os Perfis Desconexos, compostos por combinações raras ou pouco representadas, geraram baixa Coerência Jaccard. Isso ocorre devido à forte esparsidade dos vetores One-Hot e à dependência direta do modelo em relação à distribuição dos dados, o que reduz a qualidade das recomendações para categorias pouco frequentes.

A capacidade de discriminação do sistema foi analisada por meio da Curva ROC, considerando filmes com rating igual ou superior a 7,0 como relevantes. O valor de AUC obtido foi de 0,5854, indicando que o modelo possui uma probabilidade de aproximadamente 59% de classificar corretamente um filme relevante acima de um irrelevante. Embora esse valor seja superior ao desempenho aleatório, ele revela que o sistema ainda necessita de ajustes na ponderação das características e em sua representação vetorial para melhorar a precisão e a relevância das recomendações apresentadas ao usuário. Esses resultados demonstram tanto a eficácia parcial do modelo quanto as oportunidades claras de aprimoramento para versões futuras.

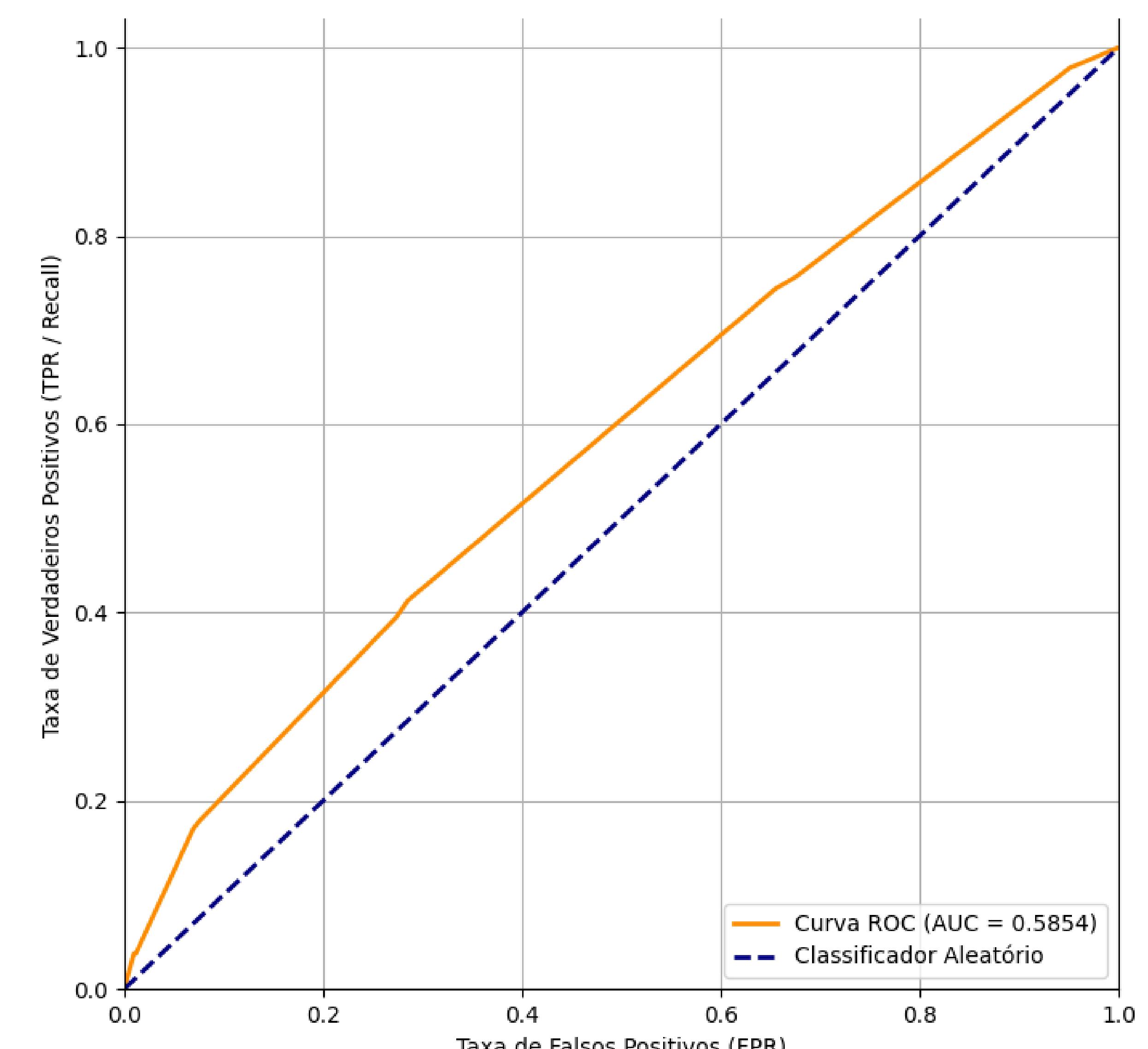
### REFERÊNCIAS BIBLIOGRÁFICAS

1. RICCI, Francesco et al. Recommender systems handbook. New York: Springer, 2015.
2. HAN, Jiawei; PEI, Jian; KAMBER, Micheline. Data Mining: Concepts and Techniques. 3. ed. Amsterdam: Elsevier, 2011.

**Tabela: Coerência Jaccard por Perfil de Entrada**

ID Perfil	Tipo de Perfil	Jaccard	% Coerência (Match)	Media Jaccard	% Coerencia
1	Perfis Comuns	0.3735	37.35%		0.2513
2	Perfis Comuns	0.3071	30.71%		25,13%
3	Perfis Comuns	0.4141	41.41%		
4	Perfis Comuns	0.4604	46.04%		
5	Perfis Comuns	0.3764	37.64%		
6	Perfis Descone	0.2267	22.67%		
7	Perfis Comuns	0.3092	30.92%		
8	Perfis Comuns	0.2765	27.65%		
9	Perfis Descone	0.2090	20.90%		
10	Perfis Descone	0.2078	20.78%		
11	Perfis Descone	0.1709	17.09%		
12	Perfis Descone	0.1997	19.97%		
13	Perfis Descone	0.1661	16.61%		
14	Perfis Comuns	0.2671	26.71%		
15	Perfis Descone	0.1270	12.70%		
16	Perfis Comuns	0.2660	26.60%		
17	Perfis Descone	0.2366	23.66%		
18	Perfis Comuns	0.2724	27.24%		
19	Perfis Descone	0.2096	20.96%		
20	Perfis Descone	0.2316	23.16%		

**Figura – Curva ROC para Avaliação de Relevância de Filmes (Rating  $\geq 7.0$ )**



### CONCLUSÃO

O trabalho desenvolveu e avaliou um Sistema de Recomendação Content-Based com ponderação de características e Similaridade de Cosseno aplicado aos dados do IMDB. Os testes com perfis sintéticos mostraram coerência nas recomendações quando as características eram frequentes, porém baixo desempenho em perfis com features raras devido à esparsidade dos dados. A Curva ROC (AUC = 0,5854) indicou capacidade moderada de discriminação entre filmes relevantes e irrelevantes, evidenciando a necessidade de ajustes nos pesos. O sistema e a interface em Streamlit configuraram uma prova de conceito funcional e base para melhorias. Como trabalhos futuros, propõem-se otimização dinâmica da ponderação, integração com métodos híbridos e avaliação de usabilidade com usuários reais.