**Sistema de Recomendação Híbrido Multiplataforma para Filmes e Séries: Integração de Conteúdo e Popularidade Temporal**

**Gabriel Guimarães de Almeida[[1]](#footnote-0)**

**Yan Victor Borba Araújo[[2]](#footnote-1)**

**Joice Bezerra Cavalcanti[[3]](#footnote-2)**

**Resumo**

Este artigo apresenta o desenvolvimento de um sistema de recomendação híbrido para filmes e séries, integrando dados de três plataformas de streaming (Netflix, Amazon Prime, Disney+). O objetivo é fornecer recomendações personalizadas e relevantes, considerando tanto similaridade de conteúdo quanto popularidade temporal. O pipeline envolve etapas de ETL, engenharia de features, implementação de algoritmo híbrido e avaliação quantitativa, com discussão sobre limitações e perspectivas futuras.

**Palavras-chave:** Sistema de recomendação; híbrido; streaming; engenharia de features; popularidade temporal; cold-start;

**Abstract**

This article presents the development of a hybrid recommendation system for movies and TV shows, integrating data from three streaming platforms: Netflix, Amazon Prime, and Disney+. The objective is to provide personalized and relevant recommendations by considering both content similarity and temporal popularity. The pipeline includes ETL processes, feature engineering, implementation of a hybrid algorithm, and quantitative evaluation, along with a discussion of limitations and future perspectives.

**Keywords:** Recommendation system; hybrid; streaming; feature engineering; temporal popularity; cold-start;

**Introdução**

A crescente oferta de conteúdo em múltiplas plataformas de streaming impõe desafios para usuários na descoberta de novos títulos. Sistemas de recomendação desempenham papel central na personalização da experiência, mas enfrentam dificuldades como o problema de cold-start e a heterogeneidade dos dados. Este trabalho propõe um sistema híbrido, combinando filtragem baseada em conteúdo e popularidade temporal, para superar tais desafios.

**Metodologia: ETL e Engenharia de Features**

O processo ETL (Extract, Transform, Load) envolveu a extração dos dados dos arquivos CSV de cada plataforma, padronização dos nomes das colunas e concatenação em um único DataFrame. A engenharia de features combinou campos relevantes (listed\_in, description, cast, director) em uma única string para vetorização textual. A vetorização foi realizada via TF-IDF, limitando a 5000 features e removendo stopwords em inglês.

# Combinação de features

def combine\_features(row):

return ' '.join(str(val) for val in [row['listed\_in'], row['description'], row['cast'], row['director']] if pd.notnull(val))

df\_merged['combined\_features'] = df\_merged.apply(combine\_features, axis=1)

**Algoritmo Híbrido: Equações e Hiperparâmetros**

O algoritmo híbrido calcula a similaridade de conteúdo entre títulos usando a matriz TF-IDF e a similaridade do cosseno. Em seguida, combina este score com um fator de popularidade temporal, modelado por decaimento exponencial:

* **Score de Popularidade Temporal:**

popularity\_score = e^{-0.5 \* (current\_year - release\_year)}

* **Score Híbrido:**

hybrid\_score = α \* content\_similarity + (1 - α) \* popularity\_score

Onde α é um hiperparâmetro de balanceamento (fixado em 0.7 após experimentação). O sistema permite filtrar recomendações por plataforma.

**Avaliação Quantitativa**

A avaliação foi realizada em hold-out, separando 20% dos títulos para teste. Foram calculadas as métricas Precision@5 e NDCG@10:

* **Precision@5:** Proporção de títulos relevantes entre os 5 primeiros recomendados.
* **NDCG@10:** Discounted Cumulative Gain normalizado para os 10 primeiros resultados, considerando relevância binária (mesmo gênero ou diretor).

Os resultados médios foram:

* **Precision@5**: 0.62
* **NDCG@10**: 0.71

O sistema apresentou desempenho robusto para títulos com histórico, mas sofreu com cold-start (novos títulos sem metadados ricos), evidenciando a limitação da abordagem baseada em conteúdo.

**Discussão: Cold-Start e Limitações**

O cold-start permanece um desafio, especialmente para títulos recém-adicionados ou com poucos metadados. A dependência de informações textuais limita a capacidade de recomendação para estes casos. Além disso, o modelo não incorpora feedback explícito do usuário, o que restringe a personalização.

**Trabalhos Futuros**

Como aprimoramentos futuros, propõe-se:

* Integração de dados de interação do usuário (ratings, histórico de visualização).
* Uso de embeddings multimodais (texto, imagem, áudio).
* Implementação de modelos de aprendizado profundo para recomendação.
* Estratégias específicas para cold-start, como recomendação baseada em clusters de gênero ou popularidade global.

**Conclusão**

O sistema híbrido desenvolvido demonstra viabilidade e eficácia na recomendação multiplataforma, conciliando relevância de conteúdo e popularidade temporal. A abordagem pode ser expandida com dados adicionais e técnicas avançadas para superar limitações atuais, especialmente no contexto de cold-start.

**Referências**

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. **Recommender Systems Handbook**. In.: Springer. 2015. Disponível em: <https://link.springer.com/book/10.1007/978-1-4899-7637-6>

AGGARWAL, Charu C. **Recommender Systems: The Textbook**. In.: Springer. 2016. Disponível em: <https://link.springer.com/book/10.1007/978-3-319-29659-3>

SARWAR, Badrul; KARYPIS, George; KONSTAN, John; RIEDL, Joseph. **Item-based collaborative filtering recommendation algorithms**. In.: Proceedings of the 10th International Conference on World Wide Web. 2001. Disponível em: <https://dl.acm.org/doi/10.1145/371920.372071>

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **Introduction to Information Retrieval**. In.: Cambridge University Press. 2008. Disponível em: <https://nlp.stanford.edu/IR-book/>

J

ANNACH, Dietmar; ZANKER, Markus; FELFERNIG, Alexander; FRIEDRICH, Gerhard. **Recommender Systems: An Introduction**. In.: Cambridge University Press. 2010. Disponível em: <https://www.cambridge.org/9780521493369>

HE, Xiangnan et al. **Neural Collaborative Filtering**. In.: Proceedings of the 26th International Conference on World Wide Web. 2017. Disponível em: <https://dl.acm.org/doi/10.1145/3038912.3052569>

1. MBA em Ciência de Dados e IA. E-mail: gabriel.dealmeida@edu.pe.senac.br [↑](#footnote-ref-0)
2. MBA em Ciência de Dados e IA. E-mail: yan.araujo@edu.pe.senac.br [↑](#footnote-ref-1)
3. MBA em Ciência de Dados e IA. E-mail: joice.cavalcanti@edu.pe.senac.br [↑](#footnote-ref-2)