

# Sistema de recomendação de filmes utilizando o algoritmo Apriori

Érica A. de Jesus<sup>1</sup>, Laysa L. de Pinho<sup>2</sup>, Vinícius de O. Costa<sup>3</sup>

Centro Universitário de Excelência (UNEX)  
Av. Artêmia Pires Freitas, s/n - Sim, Feira de Santana - BA, 44085-370

241030737@aluno.unex.edu.br, 231031777@aluno.unex.edu.br, 231031365@aluno.unex.edu.br

**Abstract.** *This paper presents a movie recommendation system for the fictional Manga Play platform, using a custom implementation of the Apriori algorithm. The goal is to suggest movies based on users' positive ratings. The process included data unification, preprocessing, and extraction of association rules using support, confidence, and lift metrics. Results show that the model produces relevant and interpretable recommendations.*

**Resumo.** *Este artigo apresenta um sistema de recomendação de filmes para a plataforma fictícia Manga Play, utilizando o algoritmo Apriori implementado do zero. O objetivo é sugerir filmes com base nas avaliações positivas dos usuários. O processo envolveu a unificação dos dados, pré-processamento e extração de regras de associação com métricas de suporte, confiança e lift. Os resultados mostram que o modelo gera recomendações relevantes e interpretáveis.*

## 1. Introdução

O crescimento das plataformas de streaming transformou a forma como consumimos conteúdo audiovisual, tornando difícil para os usuários encontrarem produções que atendam às suas preferências. Nesse cenário, os sistemas de recomendação são essenciais para oferecer sugestões personalizadas e relevantes.

Para construir sistemas eficientes, utilizam-se algoritmos como o Apriori, que identifica padrões e relações entre itens consumidos. O objetivo deste trabalho é implementar um sistema de recomendação de filmes para a plataforma Manga Play, da Manga Rosa Inc., usando o algoritmo Apriori. A ideia é gerar recomendações com base nos filmes avaliados positivamente pelos usuários, explorando as regras extraídas do histórico coletivo de preferências.

## 2. Fundamentação Teórica

O algoritmo Apriori é uma heurística de mineração de dados voltada para a descoberta de regras de associação em grandes conjuntos de dados. Seu principal objetivo é encontrar relações frequentes entre itens em um banco de dados transacional. No contexto deste projeto, os "itens" são os filmes avaliados positivamente por usuários.

Três métricas principais são utilizadas na avaliação das regras:

- **Suporte (Support):** mede a frequência com que um item ou conjunto de itens aparece no conjunto de dados.

$$\text{suporte}(A) = \frac{\text{transações contendo } A}{\text{total de transações}}$$

- **Confiança (Confidence):** mede a probabilidade de o item B estar presente, dado que A também está.

$$\text{confiança}(A \Rightarrow B) = \frac{\text{suporte}(A \cup B)}{\text{suporte}(A)}$$

- **Lift:** mede a importância da regra em relação à independência entre A e B. Valores maiores que 1 indicam correlação positiva.

$$\text{lift}(A \Rightarrow B) = \frac{\text{confiança}(A \Rightarrow B)}{\text{suporte}(B)}$$

## 2.1 Pseudocódigo do algoritmo Apriori

1. Comece com o conjunto de itemsets de tamanho 1.
2. Repita até não haver novos conjuntos frequentes:
  - a. Gere combinações de itemsets de tamanho k a partir dos conjuntos de tamanho k-1.
  - b. Calcule o suporte de cada itemset.
  - c. Elimine os que não atingirem o suporte mínimo.
3. Gere regras de associação a partir dos conjuntos frequentes.
4. Calcule confiança e lift para cada regra.

## 3. Metodologia

### 3.1 Unificação dos dados

Inicialmente, os dados foram extraídos dos arquivos movies.csv, ratings.csv, tags.csv e links.csv. A partir do dataset ratings.csv, foram selecionadas apenas as avaliações com nota superior a 3, conforme definido pelo problema, indicando que o usuário gostou do filme.

Os dados foram agrupados por usuário, de modo a obter os filmes que cada um gostou. Isso resultou em uma estrutura onde cada transação representa um conjunto de filmes apreciados por um usuário.

### 3.2 Pré-processamento

As etapas de limpeza envolveram:

- Conversão de tipos de dados.
- Remoção de duplicatas e valores ausentes.
- Agrupamento por usuário para gerar as transações utilizadas pelo Apriori.

### 3.3 Implementação do Apriori

O algoritmo Apriori foi implementado do zero, utilizando apenas bibliotecas auxiliares como pandas e numpy para manipulação de dados. A função principal identificou itemsets frequentes com base em um suporte mínimo (por exemplo, 0.02), gerou regras de associação com confiança acima de um limiar (como 0.5), e ordenou as recomendações com base na métrica de lift.

### 3.4 Recomendação

Foram considerados dois cenários:

1. Histórico completo do usuário: recomendações baseadas em todos os filmes bem avaliados anteriormente.
2. Último filme assistido: recomendações com base apenas na última avaliação positiva registrada.

## 4. Resultados e Discussões

Após a execução do sistema, foram geradas diversas regras de associação com suporte  $\geq 0.1$  e confiança  $\geq 0.5$ . Algumas das regras identificadas incluem:

- {Pulp Fiction}  $\Rightarrow$  {Reservoir Dogs}
- {The Matrix}  $\Rightarrow$  {Inception}

Essas regras indicam que usuários que gostaram de "Pulp Fiction" têm alta chance de também gostarem de "Reservoir Dogs", o que é coerente com a proximidade temática e de estilo dos filmes.

A recomendação baseada no último filme assistido mostrou-se eficaz em situações onde o histórico era escasso, mas foi perceptível que o modelo é mais robusto quando mais dados do usuário estão disponíveis.

### Limitações observadas

- A definição do suporte mínimo influencia fortemente o número e a qualidade das regras geradas.

- O modelo assume que os gostos dos usuários são estáticos e não leva em conta a ordem temporal das visualizações.
- Filmes com baixa frequência de avaliação dificilmente aparecem nas recomendações, o que limita a descoberta de novos títulos.
- A análise completa do arquivo pode levar um tempo significativo para ser processada. Além disso, em caso de erro, a demora na execução dificulta a identificação e correção do problema, tornando os testes mais lentos e menos eficientes.

## 5. Considerações Finais

O projeto alcançou seu objetivo ao desenvolver um sistema de recomendação funcional com base no algoritmo Apriori, respeitando todas as restrições impostas, como a não utilização de bibliotecas prontas para mineração de regras. O modelo conseguiu gerar recomendações relevantes e interpretáveis, permitindo inferências diretas sobre o comportamento dos usuários.

Entre os pontos positivos estão a construção manual do algoritmo, o aprendizado prático das métricas de avaliação e a flexibilidade do modelo para diferentes cenários de recomendação. Como melhorias futuras, destaca-se a inclusão de fatores temporais, ponderação por gênero ou diretor, e a comparação com outros algoritmos de recomendação, como Collaborative Filtering e Content-Based Filtering.

## Referências Bibliográficas

**AGRAWAL, R.; SRIKANT, R.** *Fast algorithms for mining association rules*. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago de Chile. Proceedings [...]. [S.l.: s.n.], 1994.

**HAN, J.; KAMBER, M.; PEI, J.** *Data mining: concepts and techniques*. 3. ed. Waltham: Elsevier, 2011.

**TAN, P.-N. et al.** *Análise de dados e mineração de dados*. Porto Alegre: Bookman, 2018.