

Recomendação de filmes com base no histórico de um usuário

David Aaron Medeiro

Bacharelado em Ciência e Tecnologia

Universidade Federal de São Paulo

São José dos Campos, Brasil

david.aaron@unifesp.br

Raissa Barbosa dos Santos

Bacharelado em Ciência da Computação

Universidade Federal de São Paulo

São José dos Campos, Brasil

raissa.barbosa@unifesp.br

Resumo—Esse trabalho tem como objetivo contextualizar e demonstrar como uma sistema básico de recomendação de filmes utilizando inteligência artificial é feito. O sistema foi feito utilizando de regra de associação, mais especificamente usando o algoritmo de Apriori, para criar relações entre avaliadores e seus respectivos filmes assistidos. Como resultado do algoritmo, após inserir um novo usuário no dataset e suas avaliações, uma recomendação contendo 3 filmes será dada.

Palavras-chave—algoritmo, filme, recomendação

I. INTRODUÇÃO E MOTIVAÇÃO

O grande crescimento da inteligência artificial tem transformado diversos aspectos no mundo inteiro, estando ela presente em grandes nichos, como na educação, na medicina e no turismo, e até mesmo em coisas mais simples, como em sistemas de trânsito.

De acordo com Gomes (2010), a inteligência artificial, ou IA, pode ser definida como um conjunto de quatro linhas de pensamento, sendo elas: os sistemas que pensam como seres humanos, os sistemas que atuam como seres humanos, os que pensam racionalmente e os que atuam racionalmente. Todos coexistindo para fazer com que as máquinas, ao terem pensamentos racionais similares aos dos humanos, consigam, também, agir como um humano.

No contexto do entretenimento, a recomendação de filmes é uma área que se beneficiou enormemente dos avanços em IA. A personalização das experiências de consumo, como a escolha de filmes, é uma forma eficaz de engajar o público e melhorar a satisfação. Além disso, a capacidade de recomendar conteúdos relevantes não só aprimora a experiência do usuário, mas também pode influenciar positivamente o consumo de mídia.

Este trabalho tem como objetivo desenvolver um sistema de recomendação de filmes que aproveita da inteligência artificial para oferecer sugestões personalizadas com base nas preferências individuais do usuário. Através de um algoritmo de recomendação, o sistema será capaz de analisar os gostos e interesses do usuário e gerar três recomendações de filmes que correspondam ao seu perfil. Com isso, pretende-se tornar mais fácil para os usuários descobrir novos filmes que eles provavelmente irão gostar, melhorando sua experiência de entretenimento.

II. CONCEITOS IMPORTANTES

A. Machine Learning e Data Mining

O Machine Learning, também conhecido como aprendizado de máquina, é um conceito da inteligência artificial que estuda como os computadores aprendem e melhoram seu desempenho, usando como base uma série de dados que são consumidos.

O Data Mining, ou mineração de dados, é um campo que usa de modelos estatísticos, algoritmos matemáticos e métodos de Machine Learning para, através de um grande volume de dados, conseguir encontrar padrões ou relações desconhecidas. Dentre os métodos de Data Mining, se encontram as categorias de aprendizado supervisionado, onde são encontrados valores, e a de aprendizado não supervisionado, onde se encontram padrões ou relações entre os dados (Fernandes, 2017).

B. Regras de Associação e Apriori

Segundo Han (2011), regras de associação são padrões frequentes que descrevem relações entre itens em um conjunto de dados transacionais. Essas regras indicam a probabilidade de ocorrência de itens em transações baseadas em suporte e confiança.

O algoritmo Apriori é uma técnica fundamental em mineração de dados utilizada para descobrir padrões frequentes e regras de associação em grandes conjuntos de dados. Baseado na ideia de que se um item é frequente, então todos os seus subconjuntos também devem ser frequentes, o Apriori funciona identificando primeiro os itens individuais que atendem a um critério de frequência mínima. Em seguida, ele gera conjuntos de itens maiores e verifica se eles também atendem a esse critério. Esse processo continua iterativamente até que não se encontrem mais conjuntos de itens que satisfaçam o critério de frequência. As regras de associação derivadas desse processo podem ser usadas para prever quais itens são frequentemente comprados juntos, tornando o algoritmo útil para aplicações como sistemas de recomendação e análise de cestas de compras.

C. Sistemas de Recomendação

Di Fante (2021) define Sistemas de Recomendação como um algoritmo ou um conjunto de algoritmos que fornece sugestões baseadas no interesse e comportamento do usuário.

Sua principal utilidade é fornecer sugestões mais assertivas de modo a atingir o usuário com maior precisão, o que consequentemente gera economia de tempo, recursos e gera uma melhor experiência para o usuário.

Esse tipo de sistema é muito utilizado atualmente na maioria das plataformas de venda e streaming, e um exemplo de fácil entendimento é a sugestão de diversos modelos de malas com capacidade específica para um usuário que pesquisou malas de capacidade semelhante. Parece mais relevante mostrar malas com essa capacidade ao invés de mostrar malas aleatórias ou até mesmo outros tipos de bolsa. Segundo Di Fante (2021), existem, basicamente, três tipos de sistemas de recomendação:

1) *Sistemas de Recomendação Baseados em Conteúdo:*

É um tipo de sistema que leva em consideração atributos do item para gerar recomendações. Por exemplo, se a intenção é recomendar livros, serão levados em consideração o autor, gênero, número de páginas, editora e outras características que um livro tem.

Esse método torna-se vantajoso pois não necessita de uma quantidade de dados iniciais tão grande quando comparado com os demais, e também não exige o conhecimento de outros usuários, somente do alvo. Além disso, ele não sofre com um problema muito comum em sistemas de recomendação, que se resume na incapacidade de recomendar algo para um usuário do qual se possui pouca informação. Entretanto, apesar das diversas vantagens, esse tipo pode ter comportamentos não desejados, como a recomendação de itens muito semelhantes (ou quase iguais, o que faz a recomendação irrelevante) e a dependência estrita e direta dos dados existentes.

2) *Sistemas de Recomendação Baseados em Filtragem Colaborativa:* Esses sistemas levam em consideração as preferências de outros usuários para realizar recomendações e podem ser subdivididos ainda em dois grupos: os baseados em memória e os baseados em modelos.

Os baseados em memória são chamados assim, pois eles basicamente funcionam com o cálculo de uma matriz de similaridade entre os usuários, e esse resultado é armazenado em memória para ser utilizado na geração das recomendações. Já os baseados em modelos constroem modelos preditivos baseados no conjunto de avaliações dos usuários. Um dos algoritmos mais utilizados nesse caso é o Singular Value Decomposition.

3) *Sistemas de Recomendação Híbridos:* Intuitivamente, combinam as abordagens dos sistemas baseados em conteúdo e os baseados em filtragem colaborativa, utilizando tanto atributos do item como preferências de outros usuários.

III. TRABALHOS RELACIONADOS

1) *Recommendation System using Apriori Algorithm:* O artigo intitulado "Recommendation System using Apriori Algorithm" explora a aplicação do algoritmo Apriori no desenvolvimento de sistemas de recomendação. O objetivo principal é utilizar o algoritmo para identificar padrões frequentes e gerar regras de associação que podem ser usadas para sugerir itens relevantes aos usuários. O algoritmo Apriori é eficaz em encontrar itens que frequentemente aparecem juntos nas

transações, permitindo que o sistema faça recomendações personalizadas baseadas em comportamentos de compra anteriores. O artigo detalha como esse processo pode ser implementado para melhorar a precisão e relevância das sugestões oferecidas aos usuários.

Através da mineração de dados e da geração de regras de associação, o artigo demonstra que o algoritmo Apriori pode significativamente aprimorar a experiência do usuário em contextos como e-commerce e plataformas de streaming. A análise de padrões de compra e a criação de recomendações baseadas nesses padrões ajudam a oferecer sugestões mais alinhadas com as preferências individuais dos usuários. Assim, o uso do algoritmo Apriori representa uma abordagem eficiente para personalizar e otimizar sistemas de recomendação, destacando a importância da análise de dados na melhoria da experiência do usuário.

2) *A Hybrid Recommendation System for Personalized Recommendation:* O artigo "A Hybrid Recommendation System for Personalized Recommendation" explora a combinação de técnicas colaborativas e baseadas em conteúdo para criar um sistema de recomendação híbrido. O objetivo é melhorar a precisão das recomendações ao integrar métodos que consideram tanto as preferências explícitas dos usuários quanto as interações passadas. O sistema híbrido combina filtragem colaborativa, que utiliza a similaridade entre usuários para fazer sugestões, com filtragem baseada em conteúdo, que analisa as características dos itens para recomendar itens semelhantes aos que o usuário já gostou.

O estudo demonstra que a combinação dessas abordagens pode superar as limitações de cada método quando usado isoladamente. A filtragem colaborativa pode enfrentar problemas com a escassez de dados ou a cold start problem (problema de novos usuários ou itens), enquanto a filtragem baseada em conteúdo pode ser limitada pela falta de diversidade nas recomendações. O sistema híbrido proposto aborda essas questões, proporcionando uma personalização mais eficaz e uma experiência de recomendação mais satisfatória para os usuários. O artigo conclui que a integração de múltiplas técnicas de recomendação pode levar a melhorias significativas na qualidade das sugestões e na satisfação do usuário.

IV. OBJETIVO

Busca-se desenvolver um sistema de recomendação de filmes baseado no gosto de cada usuário, visando auxiliar na escolha de um filme de forma rápida e simples. Com isso, a recomendação será bastante focada nas preferências de cada usuário de forma sucinta.

V. METODOLOGIA EXPERIMENTAL

Para a realização deste trabalho, diversas técnicas e bibliotecas serão usadas.

Primeiramente, sobre as bibliotecas necessárias, estão:

- Pandas, que será o dataframe utilizado para dispor as informações do dataset em linhas e colunas;
- Machine Learning Extensions (mlxtend), para a implementação do algoritmo de Apriori, que é um

dos possíveis algoritmos para se fazer uma regra de associação, e que será o utilizado neste trabalho;

- Seaborn e Matplotlib, que serão utilizada para a visualização final dos dados através de gráficos;
- Algoritmo Apriori, para fazer a regra de associação entre os parâmetros utilizados no trabalho em questão;

Para executar o algoritmo, as tabelas *movie.csv* e *rating.csv*, que podem ser encontradas aqui, foram utilizadas.

A tabela *movie.csv* contém os seguintes atributos: "movieId", "title" e "genres". Com essa tabela, pudemos verificar o nome e gênero de cada filme para o resultado final que é mostrado para o usuário.

Já a tabela *rating.csv* possui "userId", "movieId", "rating" e "timestamp", porém, o "timestamp" não foi utilizado no trabalho. Essa tabela foi a utilizada para criar as relações.

A princípio, foi plotado o gráfico abaixo para uma primeira visualização de como cada item, nesse caso cada filme, está relacionado com os usuários. Com ele, foi possível observar que bastante usuários fizeram bastante avaliações, o que resulta em um dataset bem rico para usar regra de associação.

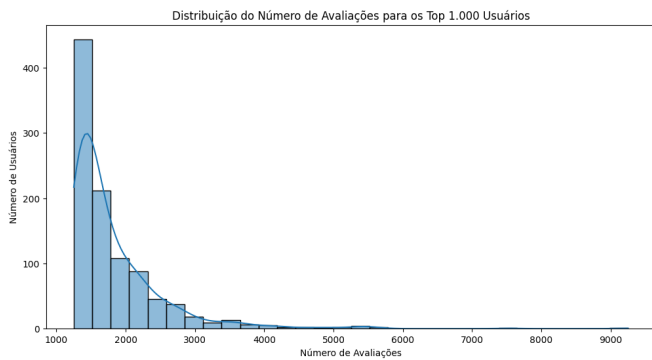


Fig. 1. Quantidade de Avaliações por Usuário

Inicialmente, as notas foram categorizadas em intervalos e com isso foi possível observar quantos filmes receberam cada nota, que foram contabilizadas em intervalos, no *rating* de cada filme.

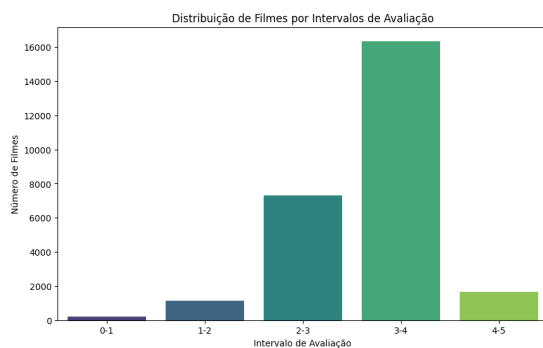


Fig. 2. Filmes e suas Avaliações

Seguindo a análise anterior, dois gráficos foram plotados, um com os 30 mais bem avaliados e outro com os 30 piores avaliados, como segue nas figuras 3 e 4.

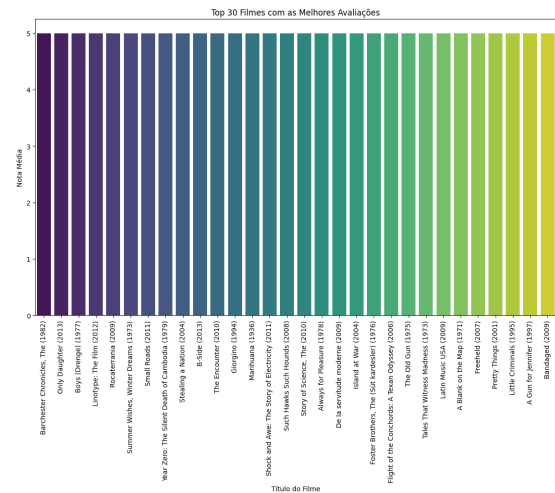


Fig. 3. Filmes com 5 estrelas

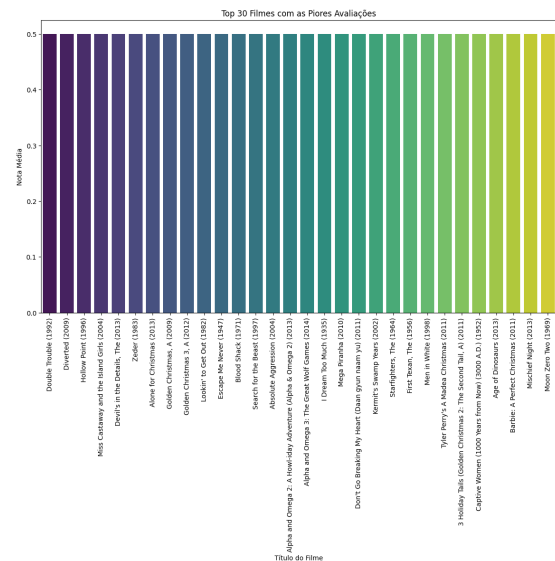


Fig. 4. Filmes com 1 estrela ou menos

Inicialmente, foram carregados ambos os conjuntos de dados, e logo após foram coletados, pelo input, quais são os três filmes, dentre os mais avaliados, que mais apeteçam ao usuário. Com esses três filmes, houve a possibilidade de adicionar três instâncias no dataframe de avaliações, como se um novo usuário os tivesse avaliado.

Após isso, foram geradas a matriz de transações, os itemsets (com suporte mínimo de 0.3) e as regras de associação (com confiança mínima da 0.5). Assim, foi possível por meio da filtragem das regras de associação (mantendo apenas aquelas nas quais os filmes avaliados pelo novo usuário aparece como antecedente) gerar as recomendações acessando os consequentes.

VI. RESULTADOS E DISCUSSÕES

Inicialmente, relacionamos os 20 filmes mais avaliados segundo a base de dados utilizada:

TABLE I
FILMES MAIS AVALIADOS

Nome

Toy Story
Twelve Monkeys (a.k.a. 12 Monkeys)
Seven (a.k.a. Se7en)
Usual Suspects, The
Braveheart
Apollo 13
Star Wars: Episode IV - A New Hope
Pulp Fiction
Shawshank Redemption, The
Forrest Gump
Speed
True Lies
Fugitive, The
Jurassic Park
Schindler's List
Aladdin
Terminator 2: Judgment Day
Dances with Wolves
Batman
Silence of the Lambs, The
 Fargo
Independence Day (a.k.a. ID4)
Godfather, The
Star Wars: Episode V - The Empire Strikes Back
Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark)
Star Wars: Episode VI - Return of the Jedi
Back to the Future
Matrix, The
American Beauty
Fight Club

Utilizando o modelo treinado com o algoritmo Apriori, foram geradas recomendações de filmes baseadas em padrões de associação encontrados nos dados. Para o primeiro teste, selecionaram-se os filmes:

- *The Matrix* (1999)
- *Back to the Future* (1985)
- *Apollo 13* (1995)

Com base nessas preferências, o sistema recomendou *The Lord of the Rings: The Fellowship of the Ring* (2001), *Toy Story* (1995) e *Independence Day* (a.k.a. *ID4*) (1996) como sugestões de filmes que o usuário provavelmente apreciaria.

Com uma pequena análise, pode-se perceber que os filmes escolhidos para o primeiro teste estão relacionados com temas envolvendo futuro, tecnologia e espaço e ciência no geral, que, geralmente, um mesmo público alvo costuma assistir. Os recomendados, por sua vez, estão bastante relacionados, pois dois dos três filmes são igualmente vistos por esse mesmo público alvo.

No segundo teste, os seguintes filmes foram escolhidos:

- *Fight Club* (1999)
- *Pulp Fiction* (1994)
- *The Godfather* (1972)

As recomendações resultantes foram *Star Wars: Episode IV - A New Hope* (1977), *The Godfather: Part II* (1974) e *The Fugitive* (1993).

Já para esse segundo exemplo, os filmes escolhidos não são tão similares, porém como o algoritmo leva em consideração os filmes assistidos por cada usuário e cria as relações entre

usuário-filme, a similaridade entre os filmes não é uma métrica que estará presente em todas as indicações.

Um ponto importante a ressaltar, é que os filmes recomendados não serão necessariamente similares com os filmes usados como entrada. Isso porque as relações são criadas baseadas nos filmes que as pessoas já assistiram, não características específicas sobre cada filme, como, por exemplo, gênero ou ano de produção.

VII. CONCLUSÕES E TRABALHOS FUTUROS

Os resultados demonstram a eficácia do algoritmo Apriori na identificação de associações entre filmes que são frequentemente apreciados em conjunto. As recomendações geradas pelo sistema são consistentes com as expectativas, apresentando filmes que compartilham características semelhantes ou estão relacionados aos filmes iniciais selecionados pelos usuários. Isso evidencia a capacidade do algoritmo de encontrar padrões significativos e gerar sugestões que estão alinhadas com os gostos dos usuários.

Além disso, é importante notar também que, embora o modelo seja eficaz em recomendar filmes com base em associações frequentes, ele pode apresentar algumas limitações, como a dificuldade em lidar com novos usuários (o problema do *cold start*, que é quando um usuário que não contém histórico no dataset utilizado gostaria de receber uma recomendação) ou a recomendação de filmes que, embora sejam populares, podem não refletir totalmente os gostos individuais dos usuários. Para superar essas limitações, futuros trabalhos podem explorar a combinação de técnicas adicionais de recomendação, como filtragem colaborativa e sistemas baseados em conteúdo, para melhorar a personalização e a precisão das recomendações oferecidas.

REFERENCES

- [1] A. P. dos Santos Porto, V. N. França Júnior, and E. Pozzebon, "Contribuições da inteligência artificial para o turismo," *Revista Rosa dos Ventos-Turismo e Hospitalidade*, vol. 15, no. 4, 2023
- [2] D. dos S. Gomes, "Inteligência Artificial: conceitos e aplicações," *Revista Olhar Científico*, vol. 1, no. 2, pp. 234–246, 2010.
- [3] J. Han, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, Massachusetts: Morgan Kaufmann Publishers, 2011, 703 p.
- [4] A. L. Di Fante, "Entendendo Sistemas de Recomendação," *Medium*, 2021.
- [5] J. Noble, "What is the Apriori algorithm?" *IBM*, 2024.
- [6] A. M. D. Rocha Fernandes and M. P. de Freitas, "Sistema de Recomendações Turísticas Utilizando Raciocínio Baseado em Casos e Geolocalização," *Brazilian Journal of Development*, vol. 7, no. 4, pp. 33752–33780, 2021.
- [7] P. M. M. Fernandes, "Aplicação de técnicas de Data Mining na Cibersegurança e Ciberdefesa—uma breve revisão," 2024.
- [8] A. S. B. L. Noronha, "Recommendation System using Apriori Algorithm," *Int. J. Sci. Res. Dev.*, vol. 3, no. 11, pp. 586–590, 2015.
- [9] K. Y. Lee, K. H. Choi, and K. W. Lee, "A Hybrid Recommendation System for Personalized Recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1530–1542, July 2013.