

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—Este 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 alguns 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, no entretenimento 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.

Por outro lado, em se tratando de filmes, o serviço de *streaming*, que é a tecnologia que permite o consumo de vídeos ou músicas sem necessidade de *download*, vem crescendo de forma exponencial nos últimos anos, e isso trouxe um novo hábito de consumo da população por essas mídias de forma imediata e em grandes quantidades [10].

De acordo com a teoria *Long Tale* proposta por Anderson, quando são comparadas uma loja física, que geralmente concentra um nicho de produtos que saem bastante em suas prateleiras, e uma loja *online*, que possui os mais variados produtos, nos dias de hoje, são altas as chances das lojas *online* possuírem um faturamento maior, o que não acontecia até um certo tempo atrás [11].

Aliado a isso, apesar desse novo padrão de consumo parecer bastante promissor, por conta de se ter um número muito grande de opções para os usuários, isso acaba fazendo com que a decisão do que comprar ou assistir se torne mais difícil. E, no caso dos *streamings*, após passar muito tempo a procura

de um filme e um ser finalmente escolhido, esse filme pode não ser do real gosto da pessoa.

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 ajuda na 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.

Nesse sentido, o atual trabalho visa conseguir, por meio da inteligência artificial, fazer uma série de recomendações de filmes para os usuários, de forma que seu histórico de filmes assistidos garantam uma recomendação útil e condizente com seus gostos.

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

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 algumas 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, em uma forma bem sucinta, o que vai melhorar sua experiência de entretenimento.

V. METODOLOGIA EXPERIMENTAL

Para a realização deste trabalho, diversas técnicas e bibliotecas foram usadas. Dentre elas, estão:

- Pandas, que será o dataframe utilizado para dispor as informações do dataset em linhas e colunas;
- Machine Learning Extensions (mlxtend), para criar a *TransactionEncoder*, que transforma a lista de transações em uma matriz de transações binárias (o que é necessário para o algoritmo escolhido, e também para a implementação do próprio algoritmo de Apriori);
- Seaborn e Matplotlib, que serão utilizadas 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*, que inicialmente começa com 27.278 linhas, contém os seguintes atributos: "movieId", "title" e "genres". Com essa tabela, foi possível verificar o nome do filme para o resultado final que é mostrado para o usuário.

Já a tabela *rating.csv*, com suas 20.000.263 linhas, possui "userId", "movieId", "rating" e "timestamp", porém, o "timestamp" não foi utilizado no trabalho. Essa tabela foi utilizada para criar as relações propriamente ditas.

A primeira coisa feita foi a verificação quanto à qualidade dos dados, então, foram verificados os seguintes pontos: se existiam valores nulos, quais os tipos de dados que seriam trabalhados, e também o formato dos *dataframes*. Depois, foi observado como os usuários mais ativos na tabela de *rating.csv* faziam suas avaliações, partindo do ponto das notas. Tudo estava de acordo, então a forma de visualização dos dados a partir disso foi através de gráficos.

A princípio, foi plotado o gráfico da Figura 1 para uma primeira visualização de como cada item, nesse caso cada filme, está relacionado com os usuários. Com ele, atrelado a consulta anterior, foi possível observar que bastante usuários fizeram bastante avaliações, o que resulta em um dataset bem rico para usar regra de associação.

Inicialmente, as notas foram categorizadas de acordo com sua avaliação e com isso foi possível observar quantas vezes cada nota foi recebida no *rating* dos filmes.

Depois, foi verificado se haviam filmes com poucas avaliações (Figura 3) e, por um gráfico utilizando o número

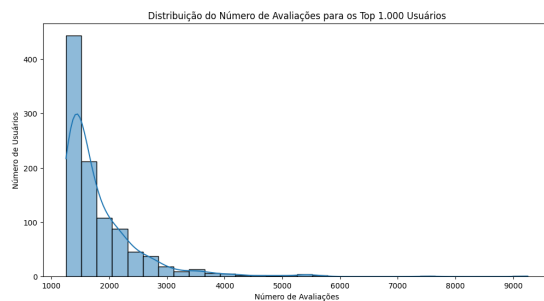


Fig. 1. Quantidade de Avaliações por Usuário

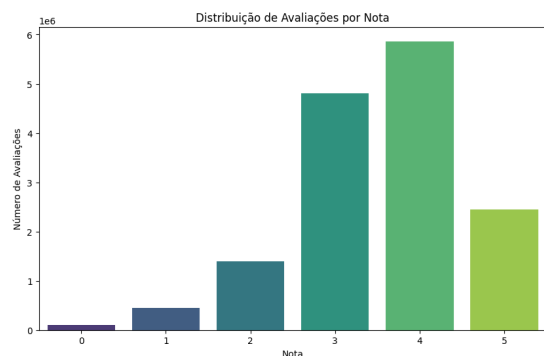


Fig. 2. Filmes e suas Avaliações

100 como parâmetro, foi percebido que bastante filme tinha pouca avaliação, o que não ajudaria muito para o resultado final nas condições do trabalho atual.

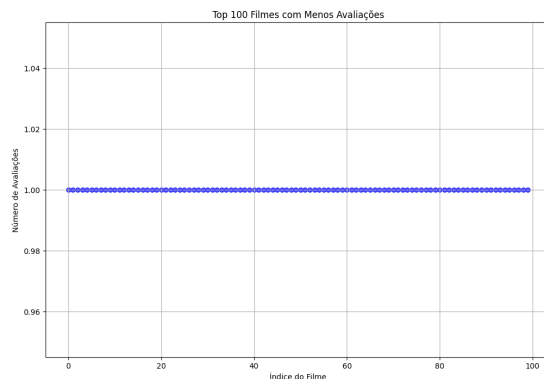


Fig. 3. Filmes e suas Avaliações

Também, foi plotado um gráfico para analisar a variância entre as avaliações, então, foi notado que alguns filmes tinham uma variância alta.

Após ambos os conjuntos de dados serem carregados e passarem pela análise inicial de pré-processamento, pelo fato de que a tabela *rating* era, inicialmente, muito extensa, foi decidido que, pelo menos, um filtro deveria ser aplicado, pois caso isso não acontecesse, o resultado final poderia ficar prejudicado por conta da alta capacidade computacional que o algoritmo exige, junto com uma base tão grande de dados

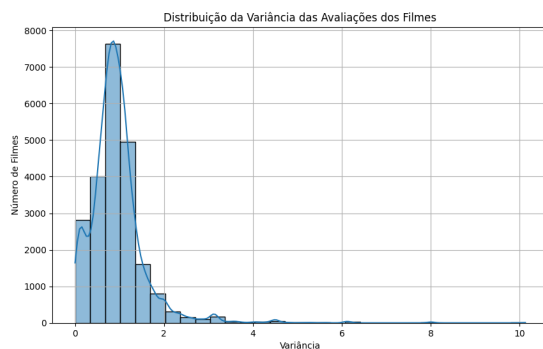


Fig. 4. Filmes e suas Avaliações

como essa.

O primeiro filtro foi relacionado com a quantidade de avaliações que um filme recebeu, que foi motivado pela Figura 3. Foi estipulado um mínimo de 20 avaliações (isso levando em consideração a quantidade de avaliações que existem), então os que não tivessem pelo menos essa quantidade de avaliações, foram excluídos. Após essa exclusão, um novo *dataframe* temporário foi usado para armazenar os filmes restantes.

O segundo, foi relacionado a variância das avaliações, com a motivadora, agora, sendo a Figura 4. Para essa análise, o novo *dataframe* "valid movies" foi utilizado, e com isso os filmes que tiveram avaliações muito diversas foram, também, excluídos.

Por fim, após a aplicação dos dois filtros, os novos *dataframes* "df movie filtered" e "df movie rating filtered" foram utilizados para o restante do código como os novos *dataframes* de base.

Depois de obter os *dataframes* atualizados, a interação com o usuário externo começou. A dinâmica é baseada nos 4 passos abaixo.

1. Perguntar a um usuário se ele deseja receber a recomendação focada em um gênero específico ou se gostaria de uma recomendação geral;
2. Filtrar (ou não) a base de rating, para que ela fique direcionada ao que o usuário gostaria;
3. Inserir alguns filmes que o usuário escolheu na base de rating;
4. Fazer a recomendação baseada nos *inputs* novos do usuário.

Três testes com configuração e entradas diferentes foram feitos e suas reflexões absorvidas.

Após cada um dos cenários serem configurados, em execuções diferentes, foram geradas as matrizes de transações, usando a métrica "lift" e o valor mínimo da métrica, ou "min threshold" como 1 ou 3.

VI. RESULTADOS E DISCUSSÕES

O primeiro teste foi sem um gênero específico e os 30 filmes mais avaliados de forma geral, segundo a base, foram impressos da seguinte maneira:

Utilizando o modelo do algoritmo Apriori, com *min threshold* igual a 1.

ID	NOME
1	Toy Story (1995)
32	Twelve Monkeys (a.k.a. 12 Monkeys) (1995)
47	Seven (a.k.a. Se7en) (1995)
50	Usual Suspects, The (1995)
110	Braveheart (1995)
150	Apollo 13 (1995)
260	Star Wars: Episode IV - A New Hope (1977)
296	Pulp Fiction (1994)
318	Shawshank Redemption, The (1994)
356	Forrest Gump (1994)
377	Speed (1994)
380	True Lies (1994)
457	Fugitive, The (1993)
480	Jurassic Park (1993)
527	Schindler's List (1993)
588	Aladdin (1992)
589	Terminator 2: Judgment Day (1991)
590	Dances with Wolves (1990)
592	Batman (1989)
593	Silence of the Lambs, The (1991)
608	Fargo (1996)
858	Godfather, The (1972)
1196	Star Wars: Episode V - The Empire Strikes Back (1980)
1198	Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)
1210	Star Wars: Episode VI - Return of the Jedi (1983)
1270	Back to the Future (1985)
2571	Matrix, The (1999)
2762	Sixth Sense, The (1999)
2858	American Beauty (1999)
2959	Fight Club (1999)

Fig. 5. Listagem dos 30 filmes sem gênero específico

- *Fight Club*
- *Sixth Sense, The (1999)*
- *Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)*
- *Godfather, The (1972)*
- *Batman (1989)*

Com base nessas preferências, o sistema recomendou *Star Wars: Episode IV - A New Hope (1977)*, *Star Wars: Episode V - The Empire Strikes Back (1980)*, *Star Wars: Episode VI - Return of the Jedi (1983)*, *Independence Day* e *Godfather: Part II, The (1974)* como sugestões de filmes que o usuário provavelmente apreciaria.

Com uma pequena análise é possível perceber que os filmes escolhidos para o primeiro teste tem forte relação com os recomendados: ambos possuem títulos contendo ação que compartilham, muitas vezes, o mesmo público alvo. É possível perceber também que uma das recomendações foi a sequência de um filme escolhido, o que certamente tem relevância para quem assistiu o primeiro.

No segundo teste, o gênero "Aventura" foi escolhido, e os filmes de possível escolha foram os da imagem abaixo.

Após a observação do usuário, foram escolhidos os seguintes filmes:

- *Star Wars: Episode IV - A New Hope*
- *Star Trek: Generations*
- *Star Wars: Episode VI - Return of the Jedi*
- *The Lord of The Rings: The Return of The King*
- *Back to the Future*

Da mesma forma que no primeiro teste, com o mesmo *min threshold* e métrica, o algoritmo foi aplicado e as recomendações foram as seguintes:

- *Toy Story*
- *Jurassic Park*
- *Mission: Impossible*

ID	NOME
1	Toy Story (1995)
10	GoldenEye (1995)
150	Apollo 13 (1995)
153	Batman Forever (1995)
260	Star Wars: Episode IV - A New Hope (1977)
316	Star Trek: Generations (1994)
329	Lion King, The (1994)
364	True Lies (1994)
480	Jurassic Park (1993)
588	Aladdin (1992)
590	Dances with Wolves (1990)
648	Mission: Impossible (1996)
733	Rock, The (1996)
1136	Monty Python and the Holy Grail (1975)
1196	Star Wars: Episode V - The Empire Strikes Back (1980)
1197	Princess Bride, The (1987)
1198	Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)
1200	Aliens (1986)
1210	Star Wars: Episode VI - Return of the Jedi (1983)
1270	Back to the Future (1985)
1291	Indiana Jones and the Last Crusade (1989)
1527	Fifth Element, The (1997)
3578	Gladiator (2000)
3793	X-Men (2000)
4306	Shrek (2001)
4993	Lord of the Rings: The Fellowship of the Ring, The (2001)
5952	Lord of the Rings: The Two Towers, The (2002)
6539	Pirates of the Caribbean: The Curse of the Black Pearl (2003)
7153	Lord of the Rings: The Return of the King, The (2003)

Fig. 6. Listagem dos 30 filmes de Aventura

- *Star Wars: Episode V*
- *Raiders of the lost Ark*
- *Indiana Jones and The Last Crusade*
- *The Lord of The Rings: The Fellowship of the Ring*
- *The Lord of The Rings: The Two Towers*

Através desse resultado, é possível perceber que os filmes estão bastante relacionados com os filmes que o usuário informou de início, pois podemos ver repetições de duas grandes sagas que foram citadas de início, que são *Star Wars* e *The Lord of The Rings*, o que demonstra uma alta chance de usuário assistir esses recomendados.

Um ponto importante a ressaltar, é que os filmes recomendados nas execuções não são necessariamente similares com os filmes usados como entrada. Isso porque as relações são criadas baseadas nos filmes que a pessoa já assistiu, não características específicas sobre cada filme, como, por exemplo, gênero ou ano de produção. Portanto, apesar de não serem todos similares, há chances do usuário assistir até mesmo os filmes que não parecem ser de seu gosto, já que outros usuários com perfis parecidos também assistiram.

Além disso, por conta do *min threshold* ser 1, que é a razão entre a confiança da regra e a expectativa de ocorrência do itemset consequente, até mesmo as relações com a razão mais baixa aparecem, recomendando, por exemplo, *Toy Story*.

No terceiro e último teste, o gênero foi o mesmo do teste 2 e as escolhas também foram as mesmas, porém o algoritmo foi treinado com o *min threshold* sendo 3. Sobre os resultados, foram os seguintes filmes:

- *Jurassic Park*
- *Star Wars: Episode V*
- *Raiders of the lost Ark*
- *The Lord of The Rings: The Fellowship of the Ring*
- *The Lord of The Rings: The Two Towers*

Como é possível observar, alguns filmes sumiram da lista de recomendação. Isso se deve ao fato de que, ao aumentar o *min threshold*, a recomendação vai ficar mais voltada aos filmes que possuem maior relação com os que estão na entrada, o que melhorou a recomendação para esse último exemplo.

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.

Por conta da alta necessidade computacional que um algoritmo mais acurado exige, alguns resultados podem não parecer muito corretos para a recomendação, então, para superar essa limitação, futuros trabalhos podem ser feitos utilizando de um computador mais potente para fazer a recomendação, o que possibilitaria a utilização da tabela de avaliações completa, antes de seus filtros. Nessa linha, além de conseguir usar as 20mi avaliações, ou até mesmo adicionar mais avaliações, e todos os filmes disponíveis, o conjunto de filmes que o usuário escolhe como entrada pode ser maior, trazendo uma abordagem mais realista.

Além de um computador com mais RAM e com GPU, a combinação de técnicas adicionais de recomendação, como o sistema de recomendação baseado em conteúdo, ou, até mesmo, o sistema de recomendação híbrido, também são ações positivas que ajudarão o algoritmo a ter uma performance melhor e é algo que pode ser explorado mais adiante.

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.
- [10] C. R. Acevedo, M. V. Navarro, P. H. V. Dignani, and B. A. Catão, "As plataformas de streaming e seu impacto no comportamento do consumidor," *REA-Revista Eletrônica de Administração*, vol. 19, no. 2, pp. 287-300, 2020.
- [11] C. Anderson, **The Long Tail: How Endless Choice Is Creating Unlimited Demand**. London: Random House Business Books, 2006.
- [12] E. Brynjolfsson, Y. J. Hu, and M. D. Smith, "From niches to riches: Anatomy of the long tail," **Sloan Management Review**, vol. 47, no. 4, pp. 67-71, 2006.