



## RECOMENDAÇÃO DE ANIMES BASEADA EM *CLUSTERING*

Sarah A. SILVA<sup>1</sup>; Carlos E. V. PEREIRA<sup>2</sup>; Diego SAQUI<sup>3</sup>

### RESUMO

A recomendação com base no agrupamento de usuários é uma maneira eficaz de recomendar novos animes aos espectadores com base em suas preferências e comportamentos semelhantes. O processo começa agrupando usuários com base em dados como gênero preferido, avaliações e histórico de visualização. Uma vez que os clusters são formados, o anime mais popular em cada cluster é identificado e recomendado aos membros do cluster. Essa abordagem leva em consideração a similaridade entre os gostos dos usuários, aumentando as chances de relevância e valorização das recomendações. Assim, tendo a capacidade de identificar animes menos conhecidos que podem atrair os usuários com base em seus padrões de consumo. Além disso, ao agrupar os usuários, é possível identificar tendências e preferências específicas dentro de cada cluster.

#### Palavras-chave:

Sistema de Recomendação; *Clusterização*; Aprendizado de Máquina; *Principal Component Analysis*; Algoritmos.

### 1. INTRODUÇÃO

A indústria de anime conquistou uma audiência crescente em todo o mundo, oferecendo uma ampla variedade de gêneros e histórias envolventes. Com cada vez mais conteúdo para escolher, os espectadores geralmente se deparam com o desafio de escolher qual anime assistir. *Clustering* é uma técnica de análise de dados que visa agrupar objetos semelhantes com base em características e medidas de similaridade (COELHO FILHO et al, 2013). No contexto da recomendação de anime, o agrupamento de usuários é usado para agrupar espectadores com perfis semelhantes considerando características como gênero de preferência, classificação de anime e histórico de exibição.

Em paralelo, a mineração de dados é uma disciplina que abrange técnicas para descoberta de padrões e conhecimento em grandes conjuntos de dados (DE AMO, 2004). Esses padrões são essenciais para identificar similaridades entre os usuários e direcionar as recomendações de animes. Então é feita a filtragem dos dados, de acordo com a opinião coletiva de usuários semelhantes (ALCANTARA e FREITAS-SÁ, 2018).

Em resumo, o objetivo do trabalho é destacar a importância dos fundamentos teóricos no desenvolvimento de sistemas de recomendação personalizados. Esses fundamentos teóricos servem

---

<sup>1</sup>Discente, IFSULDEMINAS – Campus Muzambinho. E-mail: [ssilva.srah@gmail.com](mailto:ssilva.srah@gmail.com)

<sup>2</sup>Discente, IFSULDEMINAS – Campus Muzambinho. E-mail: [kadu.kadu1303@gmail.com](mailto:kadu.kadu1303@gmail.com)

<sup>3</sup>Docente, IFSULDEMINAS – Campus Muzambinho. E-mail: [diego.saqui@muz.ifsuldeminas.edu.br](mailto:diego.saqui@muz.ifsuldeminas.edu.br)

como alicerces essenciais para criar sistemas capazes de sugerir animes de forma personalizada aos espectadores, levando em consideração suas preferências individuais e comportamentos. Em última análise, o trabalho busca ressaltar como esses princípios teóricos são cruciais para a criação de sistemas eficazes que melhorem a experiência do usuário, oferecendo recomendações mais relevantes e adequadas aos gostos e interesses de cada espectador.

## 2. MATERIAL E MÉTODOS

O primeiro passo para recomendar anime com base no *cluster* é agrupar perfis com base nas preferências do usuário e comportamentos semelhantes. Para tal, foi selecionado o dataset *Anime Recommendations Database*, que contém 500000 instâncias, e nos fornece dois csv, um com as avaliações contendo três tabelas, sendo essas: *user\_id*, *anime\_id*, *rating*. Aliado a outro csv com dados referentes aos animes: *anime\_id*, *name*, *genre*, *type*, *episodes*, *rating*, *members*. Assim, aproveitamos essas informações para criar o algoritmo de recomendação mostrado no Quadro 1.

**Quadro 1:** Algoritmo de Recomendação

```
# Passo 1: Filtrar as avaliações dos usuários com base nos gêneros mais vistos do usuário
    filtered_ratings = rating[rating['genre'].isin(user_top_genres)]
# Passo 2: Excluir animes para os quais o usuário já deu uma nota
    filtered_ratings = filtered_ratings[~filtered_ratings['anime_id'].isin(user_ratings['anime_id'])]
# Passo 3: Calcular a média das avaliações por anime
    mean_ratings = filtered_ratings.groupby(['anime_id', 'title'])['score'].mean()
# Passo 4: Ordenar as médias das avaliações em ordem decrescente
    recommended_anime = mean_ratings.sort_values(ascending=False).head(10)
```

**Fonte:** Autoral(2023)

### Métricas:

Métricas comuns de similaridade incluem a distância euclidiana e a similaridade de cosseno, que são utilizadas para calcular a similaridade entre os perfis dos usuários e direcionar a formação dos *clusters*, são elas:

- Média das avaliações por anime;
- Ordenação das médias das avaliações;
- Top 10 animes recomendados.

A aplicação de algoritmos de clustering como k-means permite dividir os usuários em diferentes clusters. Cada cluster representa um grupo de usuários com características e gostos semelhantes. O resultado é a formação de grupos heterogêneos, cada um representando um perfil específico de público.

### ***K-means:***

*K-Means* é um algoritmo de *clustering* (ou agrupamento) fornecido na biblioteca *Scikit-Learn*, que é

amplamente utilizado em análise de dados e mineração de dados. Ele encontra padrões ou estruturas subjacentes em conjuntos de dados não rotulados. Seu principal objetivo é agrupar os dados em clusters para que os pontos dentro do mesmo cluster sejam o mais semelhantes possível entre si. No contexto deste trabalho o *K-means* foi aplicado para agrupar os usuários.

### **Principal Component Analysis:**

O *Principal Component Analysis* (PCA), ou Análise de Componentes Principais, é uma técnica de redução de dimensionalidade que busca transformar um conjunto de variáveis correlacionadas em um novo conjunto não correlacionadas chamadas de componentes principais. Esses componentes capturam a maior parte da variação nos dados originais.

## **3. RESULTADOS E DISCUSSÃO**

### **3.1 Resultados**

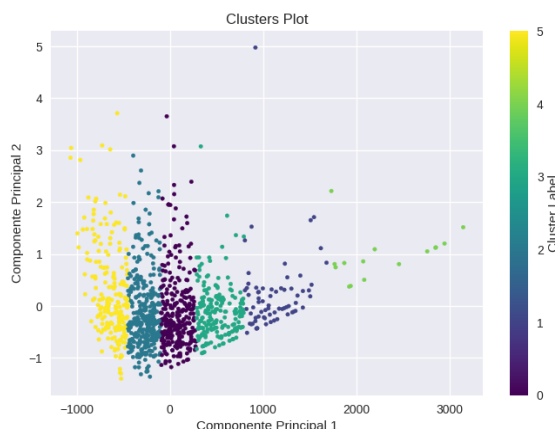
Os resultados mostraram que a recomendação de animes com base no *clustering* de usuários foi capaz de identificar animes menos conhecidos que poderiam atrair os usuários com base em seus padrões de consumo. Ademais, ao agrupar os usuários, foi possível identificar tendências e preferências específicas dentro de cada cluster. Isso proporcionou uma compreensão mais aprofundada dos gostos e interesses dos usuários, permitindo recomendações mais precisas e personalizadas como mostrado no Quadro 2 e no gráfico na Figura 1.

**Quadro 2:** Resultados de algoritmo

anime_id	title	
263	Hajime no Ippo	8.80
5258	Hajime no Ippo: New Challenger	8.73
19647	Hajime no Ippo: Rising	8.66
627	Major S1	8.38
18179	Yowamushi Pedal	8.11
31783	Yowamushi Pedal: New Generation	7.87
3090	Aa! Megami-sama!: Tatakau Tsubasa	7.67
880	Aa! Megami-sama!: Sorezore no Tsubasa	7.66
304	Aa! Megami-sama! Movie	7.63
2198	Aa! Megami-sama!: Sorezore no Tsubasa Specials	7.61

Fonte: Autoral(2023)

**Figura 1:** Plot de resultados



Fonte: Autoral(2023)

Em conseguinte foi feita a análise manual dos grupos gerados pelo K-Means, onde o número de clusters definido no K-Means foram 6 grupos de usuários, sendo estes agrupados pela média de avaliação por gêneros, expresso na planilha<sup>4</sup>.

### 3.2 Discussão

Embora a abordagem de recomendação com base no clustering de usuários tenha se mostrado promissora, é importante destacar que existem desafios e limitações. Um deles é garantir a qualidade e a quantidade adequada de dados dos usuários para um agrupamento eficiente. Além disso, é necessário considerar a escalabilidade do sistema, especialmente quando se lida com conjuntos de dados grandes.

## 4. CONCLUSÃO

Em síntese, a recomendação de animes baseada em clustering de usuários oferece uma abordagem personalizada e eficiente para sugerir novos animes aos espectadores. Ao agrupar os usuários com base em suas preferências e comportamentos semelhantes, é possível identificar animes relevantes para cada cluster e fornecer recomendações diversificadas e adaptadas aos gostos individuais. Embora essa abordagem apresenta desafios e limitações, ela proporciona uma experiência de consumo mais satisfatória, e pode se tornar uma ferramenta ainda mais precisa.

## REFERÊNCIAS

ALCANTARA, Marcela de; FREITAS-SÁ, Daniela De Grandi Castro. Metodologias sensoriais descritivas mais rápidas e versáteis, uma atualidade na ciência sensorial. *Brazilian Journal of Food Technology*, v. 21, p. e2016179, 2018.

CLEVES, V. Anime recommendation based on user clustering. Disponível em: <https://www.kaggle.com/code/viniciuscleves/anime-recomendation-based-on-user-clustering>.

COELHO FILHO, Osires Pires; MARTINHON, Carlos Alberto; CABRAL, Lucídio dos Anjos Formiga. Uma nova abordagem híbrida do algoritmo de Otimização por Enxame de Partículas com Busca Local Iterada para o problema de Clusterização de Dados. *Journal of Chemical Information and Modeling*, v. 53, n. 9, p. 1689-1699, 2013..

DBDMOBILE. MyAnimeList Dataset. Disponível em: [https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset?select=final\\_animedataset.csv](https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset?select=final_animedataset.csv).

DE AMO, Sandra. Técnicas de mineração de dados. *Jornada de Atualização em Informática*, p. 26, 2004.

---

<sup>4</sup> Planilha disponível em: [https://docs.google.com/spreadsheets/d/1ZoF9MrijQWEOtIBBM-Dkk1N6XCloWj\\_4Z6chmzIG1rl/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1ZoF9MrijQWEOtIBBM-Dkk1N6XCloWj_4Z6chmzIG1rl/edit?usp=sharing)