

# Kori

Pedro Henrique T. A.

Outubro 2025

## 1 Introdução

O tema deste trabalho é a construção de um sistema de recomendação de animes, que tem como objetivo auxiliar fãs e interessados a descobrirem novos títulos a partir de suas preferências. O sistema utiliza técnicas de processamento de linguagem natural para analisar descrições de animes e calcular a similaridade entre eles, recomendando aqueles que mais se assemelham ao escolhido pelo usuário.

Para isso, é empregado o algoritmo TF-IDF (Term Frequency-Inverse Document Frequency), que transforma as sinopses dos animes em vetores numéricos, permitindo comparar o conteúdo de forma matemática. A similaridade entre os animes é então calculada utilizando o cosseno entre esses vetores, identificando quais títulos possuem maior proximidade temática e textual.

Todo o código foi escrito em Python, utilizando bibliotecas como Pandas, Colorama, NumPy e NLTK. Os objetivos principais são:

- Fornecer recomendações de animes similares a partir de um anime escolhido pelo usuário.
- Aplicar conhecimentos da disciplina de Álgebra Linear do curso de Ciência de Dados da FATEC - Rubens Lara, ministrados pelo professor Alexandre Garcia de Oliveira, demonstrando aplicações reais de algoritmos de similaridade e espaços vetoriais.

## 2 Funcionamento

### Etapas do Sistema

O funcionamento do sistema segue as etapas abaixo:

1. O usuário pesquisa pelo nome do anime desejado e seleciona um título dentre os resultados.
2. São gerados tokens (palavras relevantes) tanto da sinopse do anime escolhido quanto dos resumos dos episódios, enriquecendo a representação textual desse anime.

3. Para cada anime do dataset, é criado um documento individual, também baseado em sua sinopse e informações relevantes.
4. O sistema utiliza o TF-IDF para transformar todos esses textos em vetores numéricos.
5. Por fim, usando a similaridade dos cossenos, compara o vetor do anime escolhido com os vetores dos demais animes do dataset (query), identificando aqueles com maior similaridade para recomendar ao usuário.

## Análise dos Ângulos entre Vetores

No sistema de recomendação, cada anime é representado por um vetor numérico gerado a partir de informações relevantes usando o algoritmo TF-IDF. Para medir a similaridade entre dois animes, calcula-se o ângulo entre seus vetores no espaço multidimensional. O cosseno do ângulo ( $\cos(\theta)$ ) entre dois vetores  $\vec{A}$  e  $\vec{B}$  é dado por:

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$

- Se o ângulo é pequeno ( $\cos(\theta)$  próximo de 1), os vetores são próximos e os animes são muito similares.
- Se o ângulo é grande ( $\cos(\theta)$  próximo de 0), os vetores são distantes e os animes são pouco similares.

Assim, o sistema recomenda os animes cujos vetores formam os menores ângulos com o vetor do anime escolhido pelo usuário, ou seja, aqueles com maior similaridade de conteúdo.

## Classificação e Visualização das Recomendações

Após calcular as similaridades entre o anime escolhido e os demais do dataset, o sistema classifica e exibe os resultados de acordo com o nível de similaridade:

- **Alta** ( $\cos(\theta) \geq 0.1$ ): representada por três círculos verdes (● ● ●) e uma barra totalmente preenchida, indicando forte similaridade (acima de 10%).
- **Média** ( $0.08 \leq \cos(\theta) < 0.1$ ): representada por dois círculos amarelos e um vazio (● ● ○) e uma barra parcialmente preenchida, indicando similaridade intermediária (entre 8% e 10%).
- **Baixa** ( $\cos(\theta) < 0.08$ ): representada por um círculo vermelho e dois vazios (● ○ ○) e uma barra pouco preenchida, indicando baixa similaridade (abaixo de 8%).

Além disso, para cada recomendação, o sistema exibe:

- O ranking da recomendação (#1, #2, ...).

- O título do anime recomendado.
- O nível de similaridade (alta, média ou baixa), o valor percentual e uma barra visual.
- Um trecho da sinopse do anime recomendado para facilitar a avaliação pelo usuário.

Essa visualização torna a experiência mais intuitiva, permitindo ao usuário entender rapidamente o grau de similaridade entre os animes recomendados e o anime escolhido.

### 3 Estrutura do Projeto

#### Descrição do Dataset

O dataset utilizado pelo sistema está localizado em `src/constants/dataset.csv` e contém informações detalhadas sobre diversos animes. Cada linha representa um anime e as principais colunas utilizadas pelo sistema são:

- **anime\_id**: Identificador único do anime.
- **Name**: Nome principal do anime.
- **synopsis**: Sinopse/descrição do enredo.

#### Descrição dos Arquivos e Diretórios

O projeto é composto pelos seguintes arquivos e diretórios principais:

- **main.py**: Script principal para execução do sistema de recomendação.
- **pyproject.toml**: Arquivo de configuração e dependências do projeto Python.
- **docs/**: Pasta de documentação do projeto (contém este arquivo .tex e outros documentos relacionados).
- **src/**: Código-fonte principal do sistema, contendo:
  - **api/**: Módulos para integração com APIs externas (AniList, AniZip).
  - **constants/**: Arquivos de dados e utilitários.
  - **searcher/**: Funções de busca e pré-processamento de texto.
  - **transformer/**: Implementação dos algoritmos de transformação e similaridade (TF-IDF, etc).
  - **utils/**: Utilitários diversos, como manipulação de caminhos e proxies.

## 4 Como rodar o projeto

Para rodar o projeto e instalar as dependências, siga os passos abaixo:

1. Certifique-se de ter o Python 3.10 ou superior instalado em seu sistema.
2. Abra o terminal na raiz do projeto (onde está o arquivo `pyproject.toml`).
3. Instale as dependências do projeto com o comando:

```
pip install .
```

4. Execute o sistema de recomendação com:

```
python main.py
```

Os resultados das recomendações são exibidos no terminal, com destaques coloridos e emojis para facilitar a visualização e tornar a experiência mais agradável.

## 5 Conclusão

O sistema de recomendação Kori foi desenvolvido para sugerir animes similares ao que o usuário escolher, utilizando técnicas de processamento de linguagem natural e álgebra linear para analisar descrições e encontrar títulos com maior similaridade. Ele é útil para quem deseja descobrir novos animes de acordo com seus interesses, de forma automatizada e personalizada.