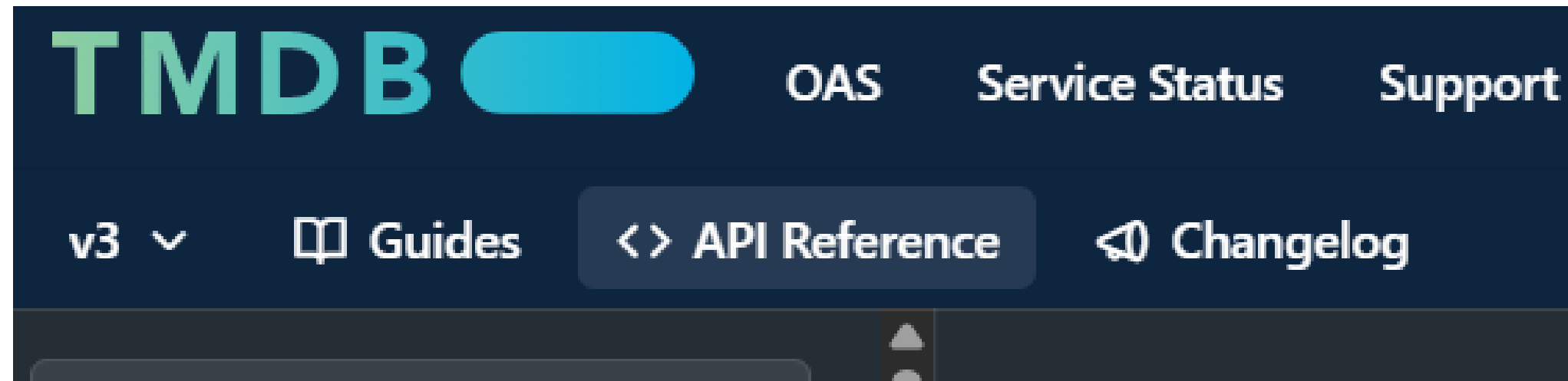


IDENTIFICAÇÃO DO GÊNERO DE UM FILME ATRAVÉS DE SUA SINOPSE UTILIZANDO NAÏVE BAYES

DANIEL NASCIMENTO LUCAS ALVES MIGUEL PEREIRA MILLENA NEVES

Banco de Dados



The Movie Database

Bibliotecas

```
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('omw-1.4')
nltk.download('wordnet')
!pip install tmdbv3api scikit-learn nltk
import re
import random
from nltk.corpus import stopwords
from tmdbv3api import TMDb, Movie, Genre, Discover
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import make_pipeline
import pandas as pd
from IPython.display import display
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import collections
from collections import defaultdict
```

Compreendendo os dados

ID

Título do filme

Sinopse

Gênero

Ano de lançamento

Popularidade

Número de palavras

```
!pip install tmdbv3api scikit-learn nltk

tmdb = TMDB()
tmdb.api_key = "e08af49bb0a5aac812535435d27024ed"
tmdb.language = "pt-BR"

movie_api = Movie()
genre_api = Genre()
discover_api = Discover()
```

Exploring creativity



Lorem ipsum
adipiscing
incididunt
Ut enim



Lorem ipsum
adipiscing
incididunt
Ut enim



Lorem ipsum
adipiscing
incididunt
Ut enim

Análise exploratória dos dados

| | | | | | | | |
|---|---------|---------------------------------------|--|-----------------------------------|------|-------------|----|
| 0 | 822119 | Capitão América: Admirável Mundo Novo | Após se encontrar com o recém-eleito presidente dos EUA, Thaddeus Ross, Sam se vê no meio de um incidente internacional. Ele deve descobrir o motivo por trás de uma conspiração global nefasta antes que o verdadeiro gênio faça o mundo inteiro ser dominado pelo vermelho. | Ação, Thriller, Ficção científica | 2025 | 1794.803500 | 45 |
| 1 | 1125899 | Cleaner | Ativistas criminosos fazem 300 reféns. Um extremista planeja assassinato em massa como mensagem para o mundo. Ex-soldada que virou limpadora de janelas tenta resgatar os reféns. | Ação, Thriller | 2025 | 1510.943375 | 26 |
| 2 | 1297763 | Batman Ninja vs. Liga da Yakuza | Nesta continuação de Batman Ninja, a Família Batman volta ao presente e descobre que o Japão desapareceu, e agora uma ilha gigante chamada "Hinomoto" flutua nos céus de Gotham City. No topo estão os Yakuza, um grupo de indivíduos superpoderosos que reinam sem honra ou humanidade e se parecem estranhamente com a Liga da Justiça. Agora, cabe a Batman e seus aliados salvar Gotham! | Animação, Ação | 2025 | 1222.743125 | 64 |
| 3 | 777443 | The Electric State | Uma adolescente órfã pega a estrada com um robô misterioso em busca do irmão há muito tempo desaparecido, encontrando no caminho um contrabandista e seu ajudante hilário. | Ficção científica, Aventura, Ação | 2025 | 932.464625 | 27 |
| 4 | 1356039 | Contra-Ataque | Uma missão de resgate faz surgir um novo inimigo, levando o capitão Guerrero e seus soldados de elite a enfrentar uma emboscada de um grupo criminoso. | Ação, Aventura, Thriller | 2025 | 817.128250 | 26 |
| 5 | 1077782 | Old Guy | Acompanhamos um assassino de aluguel que enfrenta o final de sua e fica muito emocionado quando sua companhia o coloca de volta a ativa para treinar um recém chegado jovem da Geração Z: Wihlborg, um prodígio assassino. | Ação, Comédia | 2024 | 748.912750 | 37 |
| 6 | 1126166 | Ameaça no Ar | Um piloto de avião transporta um agente federal e um fugitivo para julgamento através das montanhas do Alasca. A tensão aumenta quando verdades são reveladas e ninguém a bordo parecem ser quem realmente são. | Ação, Thriller, Crime | 2025 | 655.685125 | 34 |

Análise exploratória dos dados

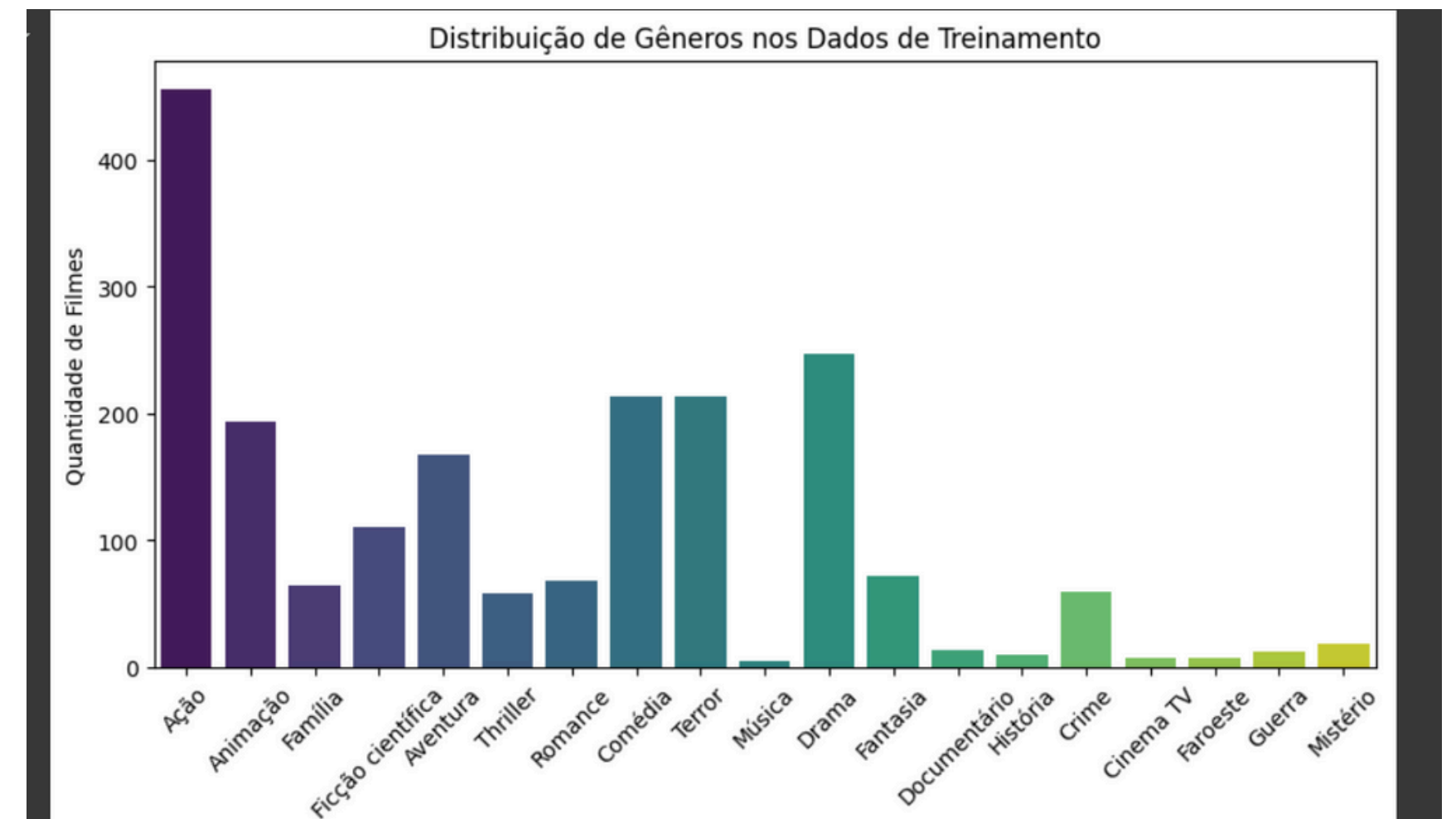
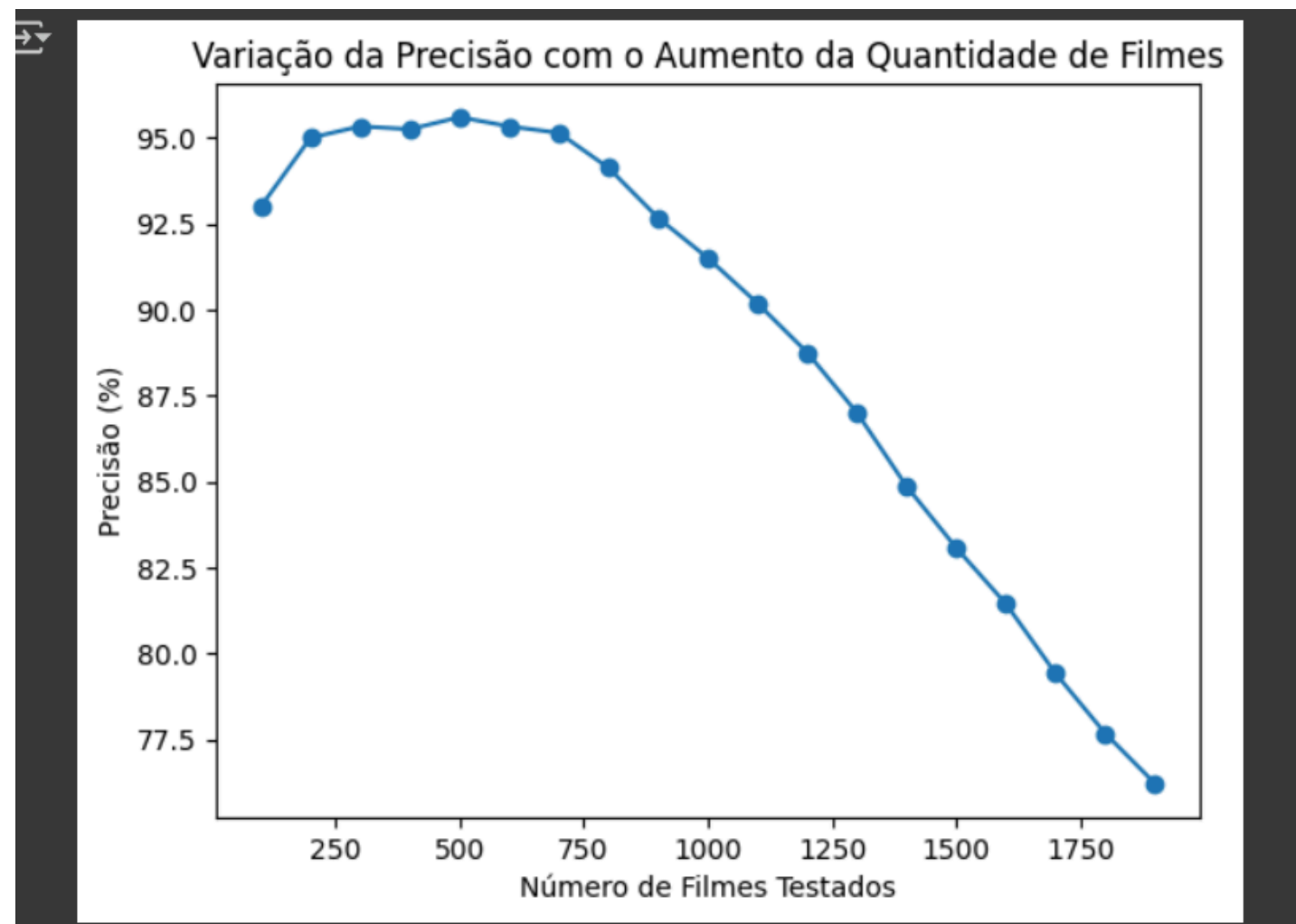
```
for genero_id, genero_nome in genero_dict.items():
    print(f"Coletando dados para o gênero: {genero_nome}...")
    contador = 0
    pagina = 1

    while contador < 200:
        filmes = discover_api.discover_movies({"with_genres": genero_id, "page": pagina})
        for filme in filmes:
            if filme.overview and contador < 200:
                generos_filme = [genero_dict[g] for g in filme.genre_ids if g in genero_dict]

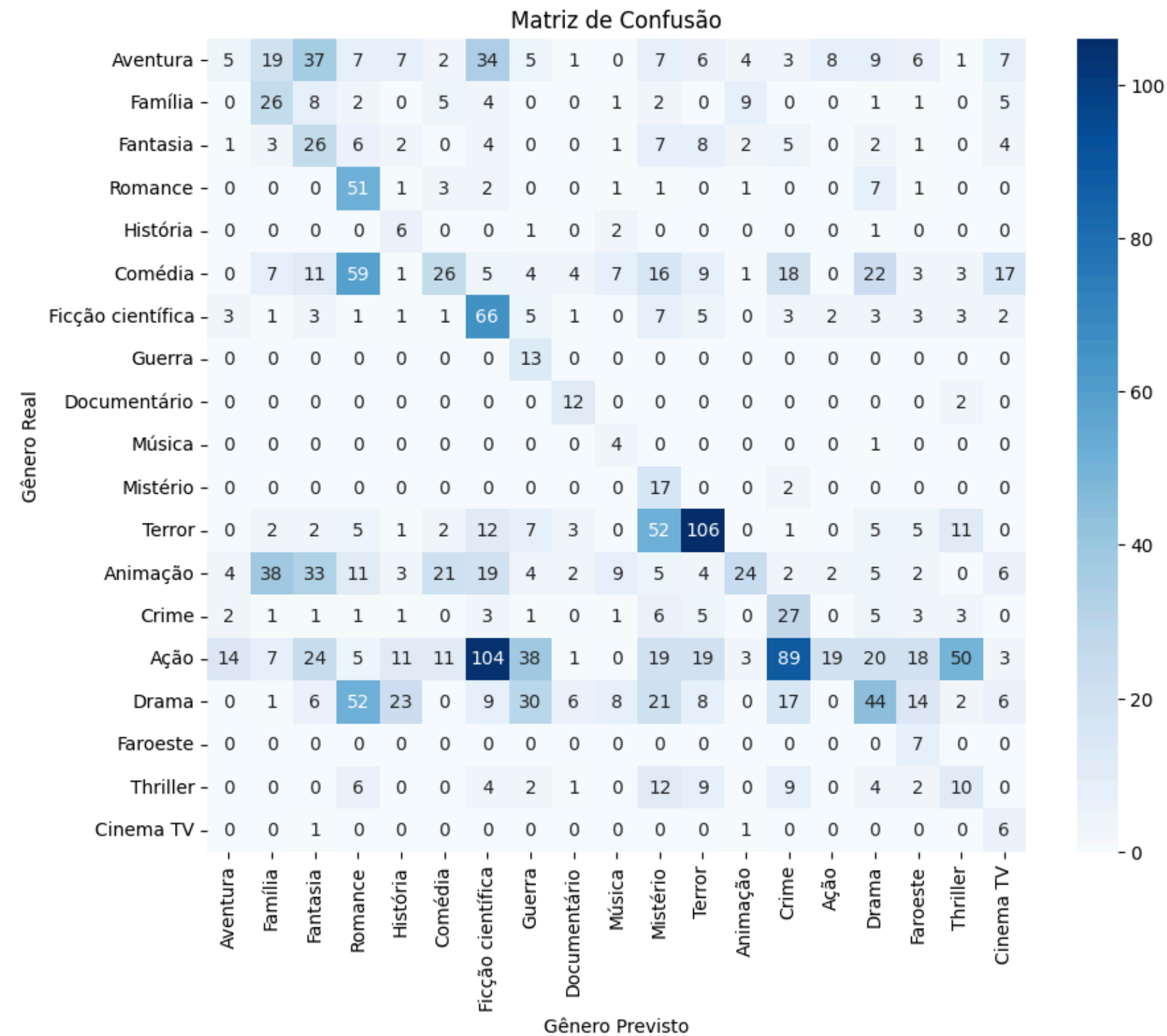
                filme_dado = {
                    "id": filme.id,
                    "titulo": filme.title,
                    "sinopse": filme.overview,
                    "generos": ", ".join(generos_filme),
                    "ano_lancamento": filme.release_date.split("-")[0] if filme.release_date else None,
                    "popularidade": filme.popularity,
                    "numero_palavras_sinopse": len(filme.overview.split())
                }
                filmes_dados.append(filme_dado)
                contador += 1
            pagina += 1

    return filmes_dados
```

Análise explanatória dos dados



Análise explanatória dos dados



Análise explanatória dos dados

| Matriz de Confusão (Gêneros Agrupados) | | | | | | | | | | | | | | | | |
|--|-------------------|----------|---------|--------------|--------|----------|--------|----------|---------|----------|------|-------|----------|---------|-----------|-------------------|
| Gênero Real | Aventura | 5 | 19 | 1 | 0 | 7 | 6 | 37 | 7 | 4 | 23 | 9 | 7 | 2 | 7 | 34 |
| | Família | 0 | 26 | 0 | 1 | 2 | 0 | 8 | 2 | 9 | 1 | 1 | 0 | 5 | 5 | 4 |
| | Documentário | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| | Música | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | Mistério | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| | Terror | 0 | 2 | 3 | 0 | 52 | 106 | 2 | 5 | 0 | 24 | 5 | 1 | 2 | 0 | 12 |
| | Fantasia | 1 | 3 | 0 | 1 | 7 | 8 | 26 | 6 | 2 | 6 | 2 | 2 | 0 | 4 | 4 |
| | Romance | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 51 | 1 | 1 | 7 | 1 | 3 | 0 | 2 |
| | Animação | 4 | 38 | 2 | 9 | 5 | 4 | 33 | 11 | 24 | 10 | 5 | 3 | 21 | 6 | 19 |
| | Ação | 16 | 8 | 2 | 1 | 37 | 33 | 25 | 12 | 3 | 291 | 29 | 12 | 11 | 3 | 111 |
| | Drama | 0 | 1 | 6 | 8 | 21 | 8 | 6 | 52 | 0 | 63 | 44 | 23 | 0 | 6 | 9 |
| | História | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 6 | 0 | 0 | 0 |
| | Comédia | 0 | 7 | 4 | 7 | 16 | 9 | 11 | 59 | 1 | 28 | 22 | 1 | 26 | 17 | 5 |
| | Cinema TV | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 6 | 0 |
| | Ficção científica | 3 | 1 | 1 | 0 | 7 | 5 | 3 | 1 | 0 | 16 | 3 | 1 | 1 | 2 | 66 |
| | | Aventura | Família | Documentário | Música | Mistério | Terror | Fantasia | Romance | Animação | Ação | Drama | História | Comédia | Cinema TV | Ficção científica |

Classificador Ingênuo de Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Resultados

