Projeto Aplicado 3

Membros:

Rafael Castro dos Santos,

Matheus Aparecido Ramos Lourenco,

Gabriel Sicari Moretti.

Professor: Murilo Gleyson Gazzola.

Sistema de Recomendação.

18/10/2023

Resumo

Este documento apresenta a primeira entrega do projeto de pesquisa científica sobre sistemas de recomendação de filmes. Nesta etapa, foi realizada a organização dos grupos de trabalho, a escolha do tema do projeto, a organização do repositório de materiais e o cronograma do projeto. Também foi definida as bibliotecas, Análise Exploratória dos Dados, treinamento do Modelo, e seus resultados.

Sumário

1. Introdução **1**
2. **Referencial teórico4**
3. **Metodologia4**

**3.1** Exploração dos dados 6

**3.2** Desenvolvimento do sistema de recomendação 6

1. **Resultado 4**
2. **Conclusão 4**
3. **Referências 4**

1 Introdução

Os sistemas de recomendação são ferramentas que utilizam algoritmos para auxiliar os usuários na busca por produtos ou serviços que possam ser do seu interesse. No caso dos filmes, os sistemas de recomendação podem ajudar os usuários a encontrar novos filmes que sejam compatíveis com seus gostos e preferências.

O objetivo deste projeto é desenvolver um sistema de recomendação de filmes baseado em dados. Para isso, será utilizado o dataset "Recomendação de Livros" disponibilizado no Kaggle. Este dataset contém informações sobre usuários, filmes e classificações de filmes.

2 Cronograma

O projeto foi feito na seguinte ordem.

1. Descrever o conjunto de dados que será utilizado.
2. Escolher o algoritmo a ser implementado.
3. Definir as bibliotecas Python necessárias.
4. Realizar uma análise exploratória dos dados.
5. Preparar e limpar os dados para o treinamento do modelo.
6. Apresentar o embasamento teórico que embasa o projeto.
7. Avaliar os resultados iniciais obtidos na etapa anterior.
8. Aperfeiçoar o fluxo de treinamento visando melhorar o desempenho do modelo.

3 Referencial teórico

Para realização deste Projeto Integrador, utilizamos a linguagem de programação Python e importamos as bibliotecas *Pandas*, *Numpy* e *Seaborn* - para a manipulação, análise e visualização dos dados -, a biblioteca *Collections* – para contar elementos em uma coleção, a biblioteca *sentence\_transformers* para tarefas de processamento de textos e cálculo de similaridade entre eles, e a biblioteca *Warnings* para desativas avisos durante a execução do código. Além disso, fizemos uso do algoritmo *NearestNeighbors* do módulo *sklearn.neighbors* que foi utilizada para o desenvolvimento dos algoritmos de recomendação.

4 Metodologia

O trabalho foi conduzido no ambiente Jupyter Notebook, escolhido por sua capacidade de executar o código célula por célula e visualizar seu resultado após a execução de cada célula. Além disso, esse ambiente oferece recursos como células de texto e imagem, permitindo uma explicação mais detalhada do código e de seus resultados. O projeto integrador foi estruturado em três fases: a exploração dos dados, o desenvolvimento do sistema de recomendação e a avaliação do sistema de recomendação.

Além disso, está sendo utilizado o GitHub como plataforma de hospedagem para o código e demais informações - <https://github.com/GabrielSicari/projeto3/tree/master>.

4.1 Exploração dos dados

Nesta fase inicial, realizamos estudos exploratórios sobre os dados presentes no dataset. O propósito foi identificar padrões e tendências que podem ser utilizados para o desenvolvimento do sistema de recomendação. Para confecção de nossos dados, utilizamos o *dataset* *The Movies Dataset* (disponível em <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>) no qual contém informações de mais de 45K filmes, 26M de avaliações feitas a partir de 270K de usuários. Dentro desse *dataset,* extraímos informações dos arquivos “movies\_metadata.csv” – que contém informações sobre os filmes como título, sinopse, data de lançamento, “ratings.csv” – avaliações dos usuários sobre determinado filme, “links.csv” - identificadores e “keywords.csv” – palavras chaves sobre o filme.

4.2 Desenvolvimento do sistema de recomendação

Na segunda etapa do projeto, concentramos nossos esforços no desenvolvimento dos algoritmos de recomendação. Inicialmente, fizemos um algoritmo utilizando apenas os dataset “movies\_metadata.csv” e “ratings.csv”, mas o resultado não foi como esperado. Como demonstra a imagem abaixo, ao solicitar uma recomendação para quem assistiu ao filme Toy Story, a menor distância que o algoritmo alcançou foi de 0.72 e recomendando um filme de terror, neste caso o “The Evil Dead”.

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Para melhorarmos a recomendação, foi preciso utilizar mais informações ao fazer a recomendação. Dessa forma, foi utilizado o dataset “keywords.csv” e a coluna de overview, de “movies\_metadata.csv”. Com isso, foi criado uma condição que verifica a similaridade entre a sinopse e palavras-chaves dos filmes. Ao realizar a consulta para recomendação do mesmo filme, podemos ter um resultado com maior precisão, com a menor distância sendo 0.309 e sendo dois filmes voltado ao mesmo público.

Texto

Descrição gerada automaticamente

Neste caso, estamos avaliando como a similaridade (menor distância) entre os filmes, quanto menor o valor, maior a similaridade entre eles. De forma simplificada, a similaridade representa o grau de semelhança entre dois objetos.

7 Referências

Dataset utilizado: https://www.kaggle.com/code/alyssonbispopereira/recomenda-o-de-filmes-ptbr/data