

Projeto de Data Mining

Movie Chat Bot

Ricardo Oliveira (PG54177) e Millena Freitas (PG54107)

Indice

1. Motivação
2. Fontes de Dados
3. Plano de Trabalho
4. Exemplos de Utilização
5. Avaliação de Desempenho
6. Objetivos Finais

Motivação

- Desenvolver um chat bot capaz de fornecer recomendações de filmes com base em várias métricas, informações sobre atores e filmes em que estes aparecem;
- Utilizar técnicas de Processamento de Linguagem Natural (NLP) e mineração de dados para compreender o *input* dos utilizadores e fornecer respostas relevantes e precisas;
- Proporcionar uma experiência interativa e familiar para os utilizadores, facilitando a descoberta de novos filmes e o acesso a informações sobre o mundo do cinema de forma mais rápida e eficiente.

Fontes de Dados

- APIs de serviços de streaming de filmes, como IMDb, TMDb (The Movie Database), etc.
- Bases de dados públicas sobre filmes e séries, contendo informações sobre o *cast*, *plot*, classificações, etc.

Plano de Trabalho

1. Análise crítica de LLM preexistentes
2. Obtenção e pré-processamento dos dados
3. Integração da LLM com as fontes de dados selecionadas
4. Desenvolvimento da interface do chat bot e a sua integração com a LLM
5. Testes e avaliação de desempenho
6. Ajustes e refinamentos à LLM, baseados nos resultados dos testes realizados

Exemplos de Utilização

Pergunta:

Indica-me três filmes dos anos 90 com avaliações acima de 9

Objetivo:

1. The Shawshank Redemption:
 - Ano: 1994
 - Duração: 2h 22m
 - Avaliação: 9.3 (2.9M)
 - Sinopse: Over the course of...
2. Schindler's List:
 - Ano: 1993
 - Duração: 3h 15m
 - Avaliação: 9.0 (1.4M)
 - Sinopse: In German-occupied Poland...
3. Ramayana: The Legend of Prince Rama:
 - Ano: 1992
 - Duração: 2h 15m
 - Avaliação: 9.2 (14K)
 - Sinopse: An anime adaptation of...

Exemplo de resposta
esperada

Exemplo Atual:

You
Indica-me três filmes dos anos 90 com avaliações a cima de 9

ChatGPT
Aqui estão três filmes dos anos 90 com avaliações acima de 9:

1. **Pulp Fiction (1994)** - Dirigido por Quentin Tarantino, este filme é aclamado pela sua narrativa não linear, diálogos afiados e elenco talentoso. No IMDb, tem uma avaliação de 8,9, mas é considerado por muitos como um dos melhores filmes dos anos 90.
2. **The Shawshank Redemption (1994)** - Baseado no conto de Stephen King e dirigido por Frank Darabont, este drama sobre esperança e redenção na prisão ganhou uma enorme base de fãs desde seu lançamento. No IMDb, tem uma classificação de 9,3.
3. **Goodfellas (1990)** - Dirigido por Martin Scorsese, este filme é uma obra-prima do gênero de máfia, baseado no livro "Wiseguy" de Nicholas Pileggi. É amplamente considerado um dos melhores filmes de Scorsese e tem uma avaliação de 8,7 no IMDb.

Fig.1: Pergunta realizada ao
ChatGPT 3.5

Correto:

IMDb ratings: 9.0 or above X Release Date: January 1, 1990 to January 1, 2000 X Movie X

Sort by Popularity

1. **The Shawshank Redemption**
1994 2h 22m R
★ 9.3 (2.9M) ☆ Rate 82 Metascore
Over the course of several years, two convicts form a friendship, seeking consolation and, eventually, redemption through basic compassion.
2. **Schindler's List**
1993 3h 15m R
★ 9.0 (1.4M) ☆ Rate 95 Metascore
In German-occupied Poland during World War II, industrialist Oskar Schindler gradually becomes concerned for his Jewish workforce after witnessing their persecution by the Nazis.
3. **Ramayana: The Legend of Prince Rama**
1993 2h 15m PG
★ 9.2 (14K) ☆ Rate
An anime adaptation of the Hindu epic the Ramayana, where Lord Ram combats the wicked king Ravana.

Fig.2: Filmes dos anos 90
com avaliação 9+ IMBD

Avaliação de Desempenho

- Testes de *benchmark* para avaliar a precisão e eficiência do chat bot;
- Comparação com outros chat bots existentes, como o ChatGPT, Blackbox.ai, etc.;
- Análise de *feedback* dos utilizadores para identificar pontos fortes e fracos.

Objetivos Finais

1. Desenvolver um chat bot funcional e preciso com conhecimento extenso sobre filmes;
2. Fornecer aos usuários uma experiência prática e intuitiva;
3. Oferecer recomendações de filmes relevantes para as pesquisas dos utilizadores;
4. Facilitar a descoberta de novos filmes e o acesso a informações sobre cinema de forma rápida e eficiente;

Projeto de Data Mining

Movie Chat Bot

Ricardo Oliveira (PG54177) e Millena Freitas (PG54107)