

ORIENTAÇÕES

Portfólio



1. TÍTULO:

BIG DATA ANALYTICS - ANÁLISE DE CLASSIFICAÇÃO DE FILMES

2. PROJETO:

Neste momento, você é um profissional Data Analysis e Big Data e sua responsabilidade será trabalhar por uma plataforma de streaming chamada "**Start-Flix**", que está buscando melhorar seu sistema de recomendação de filmes e entender o comportamento dos usuários, a partir do seu grande volume de dados. A plataforma possui uma gigante base de dados que contém milhões de avaliações de filmes, informações sobre usuários e títulos de filmes. No entanto, a empresa enfrenta grandes desafios como:

1. **Recomendações Ineficientes:** O sistema atual de recomendação não está personalizado o suficiente, resultando em baixo engajamento dos usuários.
2. **Falta de Insights sobre Preferências:** A empresa não entende completamente as preferências dos usuários por gêneros, atores ou diretores.
3. **Desempenho de Filmes:** Não há uma análise detalhada sobre quais filmes têm melhor desempenho em termos de avaliações e popularidade.

O objetivo é usar técnicas de **Data Analysis** e **Big Data** para analisar o dataset MovieLens e propor melhorias no sistema de recomendação, além de entender as preferências dos usuários e o desempenho dos filmes.

Questões para serem pensadas para resolver o problema:

1. **Quais são as principais causas do baixo engajamento dos usuários na plataforma?**

ORIENTAÇÕES

Portfólio



Resposta do Professor:

As principais causas do baixo engajamento dos usuários na plataforma **Start-Flix** estão relacionadas à falta de personalização no sistema de recomendação. Atualmente, as recomendações são genéricas e não levam em consideração as preferências individuais dos usuários. Além disso, a plataforma não utiliza dados demográficos (como idade e gênero) ou comportamentais (como histórico de avaliações) para sugerir filmes. Isso faz com que os usuários recebam recomendações irrelevantes, o que diminui a satisfação e o tempo gasto na plataforma.

2. Quem são os principais afetados por esse problema e como eles são afetados?

Resposta do Professor:

"Os principais afetados por esse problema são os **usuários da plataforma** e a **própria empresa Start-Flix**.

- **Usuários:** Eles são impactados negativamente ao receber recomendações que não correspondem aos seus gostos, o que pode levar à frustração e à redução do tempo gasto na plataforma. Isso também diminui a probabilidade de eles descobrirem novos filmes que realmente gostariam, o que poderia enriquecer sua experiência.
- **Start-Flix:** A empresa sofre com a baixa retenção de usuários e a possível migração para concorrentes que oferecem sistemas de recomendação mais eficientes. Além disso, a falta de engajamento dos usuários pode resultar em receitas menores, já que usuários insatisfeitos têm menos probabilidade de assinar planos premium ou de indicar a plataforma para outras pessoas."

3. Quais são as possíveis soluções para melhorar o sistema de recomendação e quais são os prós e contras de cada uma?

Resposta do Professor:

Existem várias soluções possíveis para melhorar o sistema de recomendação.

ORIENTAÇÕES

Portfólio



1. Filtragem Colaborativa:

- **Prós:** Essa técnica utiliza o comportamento de usuários semelhantes para fazer recomendações. Por exemplo, se dois usuários avaliaram vários filmes de forma semelhante, o sistema pode recomendar filmes que um usuário gostou para o outro. É uma abordagem eficaz para personalização.
- **Contras:** Pode sofrer com o problema do "cold start", ou seja, dificuldade em fazer recomendações para novos usuários ou filmes que ainda não foram avaliados.

2. Filtragem Baseada em Conteúdo:

- **Prós:** Recomenda filmes com base nas características dos filmes que o usuário já gostou (ex: gênero, diretor, atores). É útil para nichos específicos e não depende de dados de outros usuários.
- **Contras:** Pode ser limitada em termos de diversidade, já que tende a recomendar filmes muito semelhantes aos que o usuário já consumiu.

Uma abordagem híbrida, que combine filtragem colaborativa e baseada em conteúdo, pode ser a mais eficaz, pois aproveita os pontos fortes de ambas as técnicas."

4. Quais são as barreiras ou desafios para implementar um sistema de recomendação personalizado?

A implementação de um sistema de recomendação personalizado enfrenta vários desafios:

1. **Volume de Dados:** O sistema precisa processar milhões de avaliações e informações de usuários, o que exige infraestrutura robusta e algoritmos eficientes.
2. **Complexidade dos Algoritmos:** Técnicas como filtragem colaborativa e aprendizado de máquina exigem conhecimento técnico avançado para implementação e ajuste.

ORIENTAÇÕES

Portfólio



3. **Manutenção Contínua:** O sistema precisa ser constantemente atualizado para refletir as mudanças nas preferências dos usuários e no catálogo de filmes."

5. Como este problema se relaciona com conceitos, ideias ou questões que estudamos anteriormente em aula?

Resposta do Professor:

"Este problema está diretamente relacionado a vários conceitos que já estudamos:

1. **Análise Exploratória de Dados (EDA):** A primeira etapa do projeto envolve a análise dos dados para identificar padrões e tendências.
2. **Visualização de Dados:** A criação de dashboards é uma aplicação direta do que aprendemos sobre como apresentar dados de forma clara e visual.
3. **Big Data:** O projeto envolve o processamento de grandes volumes de dados, o que conecta diretamente com os conceitos de Big Data que estudamos, como armazenamento em nuvem e escalabilidade."

Fontes de Pesquisa Primária:

1. Livro:

- **"Programming Collective Intelligence"** por Toby Segaran.
 - Este livro oferece uma introdução prática à análise de dados e à implementação de sistemas de recomendação, incluindo técnicas de filtragem colaborativa.

2. Artigo Científico:

- **"Collaborative Filtering Recommender Systems"** por Gediminas Adomavicius e Alexander Tuzhilin.
 - Este artigo fornece uma visão geral das técnicas de filtragem colaborativa e suas aplicações em sistemas de recomendação.

3. Website:

- **Kaggle - MovieLens Dataset:**

ORIENTAÇÕES

Portfólio



- O dataset MovieLens no Kaggle é a principal fonte de dados para o projeto, contendo milhões de avaliações de filmes e informações sobre usuários.

4. Vídeo:

- "Introduction to Recommender Systems" por Andrew Ng (Coursera).
 - Este vídeo oferece uma introdução aos sistemas de recomendação e às técnicas básicas de filtragem colaborativa.

3. REQUISITOS:

Instruções para desenvolvimento e entrega do trabalho.

Etapas do Projeto:

1. Coleta e Estruturação de Dados (Google Colab):

- Importe o dataset MovieLens para o Google Colab.
- Limpe os dados.

2. Análise Exploratória (Python - Google Colab):

- Importe os dados para o Google Colab.
- Use a biblioteca Pandas para explorar os dados.
- Crie visualizações com Matplotlib e Seaborn.

3. Visualização de Dados:

- Crie um dashboard interativo com pelo menos 4 painéis:

4. Relatório Final:

- Desenvolva um relatório detalhado com base na análise dos dados.
- Inclua gráficos e tabelas gerados no Google Colab.

5. Ferramentas de apoio

- GitHub para apoio e inserção dos resultados do trabalho

ORIENTAÇÕES

Portfólio



- Google Colab para execução dos códigos Python
- Linguagem de dados: Python
- Word ou similar para preenchimento do relatório final

PASSO A PASSO DA EXECUÇÃO:

1. Crie sua conta no GitHub (sugiro usar sua conta Google), pois ela servirá para você armazenar o resultado do seu projeto de Portfólio e para conseguir acessar o Dataset (conjunto de dados da tarefa) e as instruções de comando.
2. Acesse e baixe as instruções e os dados no endereço GitHub no link abaixo:
<https://github.com/PROFSANTARELLI/BIG-DATA>
3. Siga as instruções contidas no documento de execução.
4. Colete os resultados no documento Word (ou similar) e monte o relatório de forma lógica, com introdução, demonstração dos resultados e conclusão.
5. Transforme o relatório em PDF.
6. Grave um vídeo de até 3 minutos contando sobre a execução e os resultados.
7. Envie o relatório e o vídeo para avaliação do Tutor.
8. Guarde no seu Github todos os documentos gerados nesse projeto de portfólio.

4. ENTREGÁVEL:

ORIENTAÇÕES

Portfólio



1. Relatório de Análise:

- Desenvolver relatório com os principais insights e recomendações.
- Inclui gráficos e tabelas gerados no Python e no Looker Studio.
- Faça uma conclusão final dos dados analisados e do trabalho realizado

2. Vídeo de Apresentação:

- Vídeo de até 3 minutos explicando o projeto e os resultados.

- O aluno deverá entregar o relatório em PDF e o vídeo em MP4.

ORIENTAÇÕES

Portfólio



Roteiro do estudante

1. Leia o PROJETO:

Sua primeira tarefa é se tornar um detetive. Use os recursos que fornecemos, mas também não tenha medo de sair em busca de suas próprias informações adicionais. Queremos que você compreenda o tópico e explore o problema a fundo.

2. Entenda os OBJETIVOS:

Muita atenção para não perder o foco durante o estudo. Você precisa compreender qual é o objetivo para não perder isso de vista durante todo o processo.

3. Siga os REQUISITOS:

Este é o momento de entender as regras do jogo. Os requisitos dizem como você deve desenvolver o projeto e quais são os recursos necessários para chegar ao produto final.

4. Não ignore o Exemplo:

Aqui você verá exatamente o que é esperado de você. É fundamental que você chegue a um resultado que siga essa referência.

5. Aponte a melhor solução:

Depois de reunir todas as suas informações, é hora de colocar a mão na massa. Busque identificar padrões, relações, causas e efeitos. Procure contradições ou conflitos de informação. Seja crítico e questione o que você encontrou durante sua investigação. Agora é o momento de se tornar um(a) solucionador(a) de problemas. Com base no que você descobriu e analisou, use os requisitos a seu favor e entregue a melhor versão do seu trabalho!

ORIENTAÇÕES

Portfólio



Como seu trabalho será avaliado?

Sua avaliação será baseada em uma série de critérios, com uma pontuação de até 4,0 pontos.

Lembre-se, o objetivo aqui não é apenas obter a pontuação máxima, mas aprender durante o processo. Aproveite essa jornada!