# Software Product Streaming Data Analysis

ALUNO: HIGOR NOGUEIRA SOUZA PROFISSÃO: ENGENHEIRO DE DADOS / MACHINE LEARNING MATRÍCULA:1904220

PROFESSOR: ANTÔNIO DE OLIVEIRA DIAS

MATÉRIA: SOFTWARE PRODUCT, ANALYSIS, SPECIFICATION, PROJECT E IMPLEMENTATION

PROJETO INDIVIDUAL PARA A MATÉRIA SOFTWARE PRODUCT

Projeto: Desenvolvimento de uma plataforma de streaming de vídeos que inclui análise de dados para categorização automática de títulos de filmes.

Descrição: Este projeto visa criar uma plataforma de streaming de vídeos que ofereça uma experiência personalizada aos usuários. Um dos principais recursos desta plataforma é a capacidade de categorizar automaticamente os títulos de filmes com base em informações textuais, como sinopses e metadados, permitindo uma melhor organização e recomendação de conteúdo.

Fluxo de Trabalho:

Coleta de Dados: Os dados brutos, incluindo informações sobre títulos de filmes, sinopses e outros metadados, são coletados de diversas fontes e armazenados em uma planilha csv utilizada para teste (APIs de Dados, Web Scraping, ETL)

Pré-Processamento e análise de dados: Cientistas de dados usam técnicas de processamento de linguagem natural (NLP) e aprendizado de máquina para analisar os dados textuais e classificar automaticamente os títulos de filmes em categorias relevantes, como gênero, classificação etária, etc.

Treinamento de Modelos: Modelos de machine learning são desenvolvidos e treinados no ambiente do Google Colab usando dados históricos para classificar os títulos de filmes. Bibliotecas de Aprendizado de Máquina: Scikit-Learn e processamento de Linguagem Natural (NLP): NLTK, Google Colab, Git.

Integração com a Plataforma: Os modelos treinados são integrados à plataforma de streaming, permitindo que novos títulos sejam categorizados automaticamente à medida que são adicionados à biblioteca. (Desenvolvimento Web, HTML, Flask, Banco de Dados: PostgreSQL, APIs RESTful)

Apresentação Personalizada: Com base nas categorias atribuídas automaticamente, a plataforma pode oferecer uma experiência de usuário personalizada, recomendar conteúdo relacionado e facilitar a navegação ( Collaborative Filtering, Content-Based Filtering).

<https://colab.research.google.com/drive/1kxezPd8B52jQMhc_xFVCSvazd-3PURLH?usp=sharing>

<https://trello.com/b/QsvjQUz0/requisitos-do-projeto-de-software-an%C3%A1lise-de-dados> [higor.souza@aluno.faculdadeimpacta.com.br](mailto:higor.souza@aluno.faculdadeimpacta.com.br)