



Sprint II

INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL

Integrantes:

Laura Zoia - 93294

Pedro Nogueira – 94402

Gabriel Moura - 93940

Heitor Mancini - 94288

1. Introdução

No desafio deste ano, fomos agraciados com tarefas propostas pela Sanofi, uma empresa francesa do setor farmacêutico. Os desafios abrangem desde dashboards, IA, ML, até dados e muito mais. Nosso grupo, composto por quatro membros, está entusiasmado em contribuir para mudanças que tornarão o trabalho dos colaboradores da empresa mais ágil. É evidente que os problemas apresentados representam atrasos tecnológicos significativos, e estamos muito animados em mostrar aos gestores que propuseram os desafios formas de inovar e otimizar um tempo considerável.

Em alinhamento com a iniciativa da Sanofi sobre "Análise de dados" para uma compreensão aprofundada do histórico de compras, este projeto tem como objetivo desenvolver uma ferramenta que automatize tanto a extração quanto a análise de dados, seguindo critérios previamente estabelecidos. A meta é apresentar as informações de maneira visual e acessível, facilitando a compreensão pelos usuários.

2. Objetivos e Escopo

Desenvolver uma ferramenta de análise de dados automatizada que permita a extração, limpeza, análise e visualização de dados com base em critérios preestabelecidos.

Proporcionar informações de maneira visual e de fácil compreensão para os usuários, facilitando a tomada de decisões com base nos dados analisados.

Reduzir o tempo e esforço necessário para a análise de dados, tornando o processo mais eficiente e ágil.

Permitir a personalização dos critérios de análise e das visualizações de acordo com as necessidades específicas de cada usuário ou projeto.

- A ferramenta será capaz de se integrar a diferentes fontes de dados, como bancos de dados SQL, arquivos CSV, bases em Excel, entre outros, para realizar a extração de dados.

- A equipe da Sanofi nos informou que utilizavam apenas soluções em Excel para o processo, já nos ajudando quando for extraído os dados.

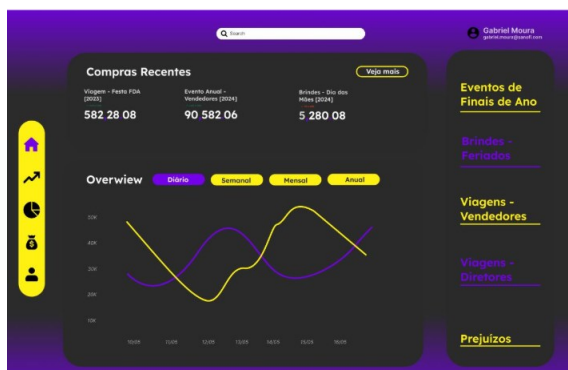
- Utilizará técnicas de análise de dados, como estatísticas descritivas, modelagem preditiva e análise de séries temporais, para extrair insights relevantes dos dados.

- Apresentará os resultados da análise por meio de visualizações interativas, como gráficos, tabelas e dashboards, que permitirão aos usuários explorarem os dados de forma intuitiva.

- A solução de Gráficos e alternativas visuais poderá ser feita em aplicações como Power BI e Tableau.

- Permitirá a automação de tarefas recorrentes, como atualização periódica dos dados e geração de relatórios automatizados.

- Será desenvolvida uma interface de usuário amigável e intuitiva, que permitirá aos usuários configurarem os critérios de análise e explorar os resultados de maneira fácil e eficiente.



3. Contextualização

O projeto será baseado prioritariamente em software, sendo assim não será necessária nenhuma forma de integração com um hardware.

Ele pode ser executado em ambientes de computação em nuvem ou localmente em servidores dedicados, assim poderá ser compartilhado para a equipe automaticamente. Também serão utilizadas linguagens como Python para desenvolvimento da lógica do sistema e Javascript para a criação de visualizações interativas.

O Python tornou-se uma escolha preferencial para desenvolvedores de IA, dada a sua capacidade de ultrapassar os limites da programação tradicional e desenvolver soluções autodidatas e adaptáveis.

Poderão ser utilizadas IDEs como Jupyter Notebook ou Google Colab para desenvolvimento em Python e Visual Studio Code para desenvolvimento em JavaScript.

Serão explorados frameworks como Pandas e NumPy para manipulação de dados em Python, além de bibliotecas de visualização como Matplotlib e Plotly.

O Pandas pode ser utilizado para várias atividades e processos, entre eles: limpeza e tratamento de dados, análise exploratória de dados (EDA), suporte em atividades de Machine Learning, consultas e queries em bancos de dados relacionais, visualização de dados, web scraping e muito mais. Além disso, também possui ótima integração com várias outras bibliotecas muito utilizadas em Ciência de Dados, tais como: Numpy, Scikit-Learn, Seaborn, Altair, Matplotlib, Plotly, Scipy e outros.

4. Materiais e Métodos

Para desenvolver a ferramenta de análise de dados automatizada, adotaremos uma abordagem metódica e estruturada que abrange desde a extração até a visualização de

dados. O projeto será dividido em várias etapas, cada uma utilizando ferramentas e técnicas específicas para garantir a eficiência e a precisão dos resultados. A seguir, detalhamos os materiais e métodos utilizados em cada etapa do projeto.

Para este projeto, serão utilizadas as seguintes tecnologias:

- 1. Engenharia, Análise e Ciência de Dados:** para o fluxo, análise e processamento dos dados;
- 2. Bot com Inteligência Artificial:** para a automação de tarefas e aprimoramento da análise de dados;
- 3. Visão Computacional:** para o desenvolvimento de visualizações interativas e intuitivas;
- 4. APIs de integração:** para conectar a ferramenta a diferentes fontes de dados e serviços externos.

CIÊNCIA DE DADOS

- **Pandas:** uma biblioteca Python amplamente utilizada para manipulação e análise de dados;
- **NumPy:** outra biblioteca Python essencial para computação numérica, útil para operações matemáticas e manipulação de arrays multidimensionais;
- **Scikit-learn:** uma biblioteca Python de aprendizado de máquina que oferece uma ampla gama de algoritmos para classificação, regressão, clustering, entre outros;
- **TensorFlow ou PyTorch:** Frameworks populares para construção e treinamento de modelos de aprendizado de máquina e redes neurais.

BOT COM IA

- **Pandas:** o framework Pandas é uma ferramenta poderosa para manipulação e análise de dados em Python e pode desempenhar um papel crucial em diversos aspectos do projeto de automação da análise de dados;
- **Scikit-learn:** além de ser usado para análise de dados, o Scikit-learn oferece uma variedade de ferramentas para automação de tarefas de aprendizado de máquina, como busca de hiperparâmetros e validação cruzada;
- **AutoML:** plataformas de Automated Machine Learning (AutoML), como Google Cloud AutoML e H2O AutoML, que automatizam o processo de construção de modelos de machine learning, desde a seleção do modelo até o ajuste dos hiperparâmetros;
- **Bot:** podemos utilizar também sites que fazem Bots de uma maneira mais simplificada, como a programação em blocos. Podemos utilizar os sites: UiPath, Blip Assist, Zenvia NLU ou Watson.

VISÃO COMPUTACIONAL

- **OpenCV:** uma biblioteca de visão computacional de código aberto amplamente utilizada, que oferece uma variedade de algoritmos para processamento de imagens, detecção de objetos, reconhecimento de padrões, entre outros;

- **Matplotlib e Seaborn:** bibliotecas Python para criação de visualizações estáticas de dados, como gráficos de barras, histogramas, dispersões, entre outros;
- **Plotly e Bokeh:** bibliotecas Python para criação de visualizações interativas e dinâmicas, como gráficos de linha interativos, mapas de calor, gráficos 3D, entre outros.

APIs DE INTEGRAÇÃO

- **Requests:** uma biblioteca Python para fazer requisições HTTP, útil para interagir com APIs web;
- **Flask ou Django:** frameworks web Python para desenvolvimento de APIs RESTful, que podem ser usados para expor funcionalidades da ferramenta para integração com outros sistemas.

5. Comentários e Perspectivas

Visamos atender as expectativas da Sanofi para um projeto de análise de dados bem-sucedido, é necessária uma combinação de tecnologias avançadas, compreensão das necessidades do cliente, abordagem holística e foco na segurança e privacidade dos dados.

Essas escolhas tecnológicas estabelecem uma base robusta para a criação de uma ferramenta que simplifica não apenas a análise de dados, mas a torna poderosa e acessível para diversos usuários. Com a integração de tecnologias avançadas, nossa ferramenta será capaz de lidar com grandes volumes de dados, extrair insights valiosos e apresentá-los de maneira visualmente atraente. Além disso, a capacidade de se conectar a diferentes fontes de dados possibilita uma abordagem completa para a análise, permitindo aos usuários explorarem informações de várias fontes. Essas decisões asseguram que nossa ferramenta seja flexível, rápida e eficaz, capacitando os usuários a fazerem escolhas informadas com confiança