

Orientação - Atividade Somativa 2:

Desenvolvendo um projeto de IA em Saúde

Agora que estamos na reta final da disciplina, e você já trabalhou com as principais técnicas de IA em Saúde, chegou a hora de comprovar, na prática, tudo que foi aprendido.

Nesta atividade, você deve desenvolver um projeto completo de IA em Saúde, focando no uso de técnicas de Machine Learning (ML) ou Processamento de Linguagem Natural (PLN). Além destas, você deve obrigatoriamente explorar técnicas de análise e visualização de dados.

Você deve **selecionar um dataset ainda não explorado durante a disciplina**, estruturado ou não-estruturado, de qualquer repositório público de dados (alguns estão listados a seguir). Caso utilize dados privados, de um use case real, lembre-se apenas das questões de privacidade e segurança dos dados. **Ao escolher o dataset, confirme com seu professor-tutor se o mesmo é adequado e atende os requisitos do projeto.**

[Kaggle](#)

[Github](#)

[Datasus](#)

[UCI Machine Learning Repository](#)

[Google Dataset Search](#)

O projeto inclui a produção dos seguintes artefatos:

1. Descrição do cenário

1. Qual é o objetivo do seu projeto?

2. Qual é o contexto da organização para o qual o projeto está sendo desenvolvido? (pode ser fictício)

3. Qual é a motivação para a construção do projeto neste cenário? Quais problemas ele busca resolver?
2. Descrição das tecnologias adotadas
 1. Qual abordagem de desenvolvimento será escolhida?
 2. Quais técnicas de PLN e ML serão utilizadas?
 3. Quais bibliotecas/pacotes/frameworks serão utilizados?
 4. Justificar o uso de todas as tecnologias adotadas.
3. Descrição do dataset escolhido
 1. Quais dados constam no dataset?
 2. Quais dados serão utilizados no desenvolvimento do projeto?
 3. Qual a origem dos dados?
 4. Em quais tarefas o dataset poderia ser útil?
4. Programa funcional (script de manipulação dos dados) – a seguir os requisitos mínimos
 1. Deve conter uma etapa de **abertura dos dados** (preferencialmente importados para um dataframe).
 2. Deve conter uma etapa de **pré-processamento dos dados**. Caso não seja realizada esta etapa, ela deve ser justificada em comentários no próprio código.
 3. Deve conter uma etapa de **análise exploratória dos dados**, incluindo ao menos 3 tipos de visualização de dados, e um descritivo de quaisquer insights gerados por esta etapa.
 4. Deve conter uma etapa de **treinamento de um modelo de ML**.
 5. Deve conter uma etapa de **predição e explicabilidade (interpretabilidade)** do modelo. Em caso de não possibilidade, justificar.
 6. Deve conter uma etapa de **avaliação do modelo**.
 7. Deve conter uma etapa de **conclusão**, que deve resumir o funcionamento do projeto, apresentar os principais achados, e descrever os próximos passos para melhoria do método (se necessários).

O código-fonte e parte escrita do projeto devem ser entregues em um único Notebook Python (arquivo ipynb). Links para o Google Colab ou outros formatos não serão aceitos.