Faculdade de Informática e Administração Paulista – FIAP

Campus Aclimação

Turma 1TDSR

Artificial Intelligence and ChatBot

Challenge – 3º Sprint

Projeto ELO - IMREA

Gabriel Nogueira Peixoto

Giovanna Neri dos Santos

Mariana Inoue

26 de setembro de 2025.

São Paulo.

**DESCRIÇÃO:**

**EXPLICAÇÃO DOS DADOS**

Escolhemos um dataset de registros médicos de consultas, que foram coletados no Centro de Especialização em Reabilitação Física e Intelectual (CER) da Universidade do Vale do Itajaí, que atende pacientes pelo SUS no sul do Brasil.

**Fonte:** <https://data.mendeley.com/datasets/wm6w2fvkfj/1>

**SIGNIFICADO DOS DADOS**

O dataset foi construído com o intuito de estudar no-show em consultas médicas, que é uma prática de ter um compromisso, nesse caso uma consulta médica agendada e não comparecer. É importante ressaltar que no-show e absenteísmo, falta ou atraso independente do motivo, são diferentes na prática, porém o dataset escolhido ainda se mantém relevante, pois o lugar de coleta dos dados (CER) tem atuação bem semelhante ao IMREA, que é a instituição que estamos atendendo no Challenge.

**QUANTIDADE E QUALIDADE**

São cerca de 49,593 registros médicos de consultas, que engloba consultas que foram realizadas e não realizadas. A amostra de consultas que tiveram o no-show pela análise prévia gira em torno de 1%, o que é relativamente baixo, mas ainda temos variáveis que não foram possíveis de encontrar em outros datasets que podem enriquecer o contexto de cada consulta.

**RÓTULOS E CARACTERÍSTICAS IMPORTANTES**

Os rótulos abaixo são as principais colunas que irão nos ajudar a selecionar e correlacionar os dados dos atendimentos para podermos ter uma boa base para o treinamento:

* **Specialty:** especialidade médica da consulta
* **Disability:** tipo de deficiência do paciente (motora ou intelectual)
* **Appointment Year / Month / Shift:** data e turno da consulta
* **No-show:** se o paciente compareceu ou não
* **No-show Reason:** motivo da ausência
* **Weather conditions:** temperatura, chuva e intensidade no dia da consulta
* **ICD (CID):** código da doença associada ao paciente

**OBJETIVO DA ANÁLISE**

O objetivo desta análise vai ser treinar um modelo que seja capaz de prever a probabilidade de um paciente faltar à consulta, fortalecendo nossa solução que tem como meta diminuir o absenteísmo, através dar um suporte melhor aos pacientes nas consultas remotas através de um tutorial de ambientação e uma agenda.