# Healthcare Appointment No-Show Prediction

## 1. Introduction

Hospitals often face the challenge of patient no-shows, which can disrupt schedules, waste resources, and delay treatment. This project aims to analyze factors that influence missed appointments and to develop a predictive model that can help medical administrators reduce the no-show rate.

## 2. Abstract

The goal of this project is to predict whether a patient will miss their scheduled healthcare appointment using historical appointment data. By analyzing features such as age, gender, medical conditions, waiting time, and SMS reminders, we built a decision tree classifier to predict no-shows. We also visualized key insights using Power BI to support data-driven operational improvements.

## 3. Tools Used

Programming: Python  
Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn  
Visualization: Power BI  
Model: Decision Tree Classifier (Sklearn)

## 4. Steps Involved in Building the Project

Step 1 – Data Collection

We used a publicly available dataset containing over 100,000 medical appointments in Brazil, with information on whether the patient showed up.

Step 2 – Data Cleaning & Preprocessing

We removed irrelevant columns (PatientID, AppointmentID), converted date formats, filtered out negative ages, encoded categorical values (Gender, No-show), and created features like WaitingDays and AppointmentWeekday.

Step 3 – Exploratory Data Analysis (EDA)

Visualizations were created to analyze relationships between no-shows and features like age, weekdays, SMS reminders, and medical conditions. Notable findings:  
- Longer waiting times increase no-show probability.  
- SMS reminders correlate with higher attendance.  
- Older adults and patients with chronic conditions show higher adherence.

Step 4 – Model Building

A Decision Tree Classifier was trained using selected features (Age, Gender, Scholarship, SMS\_received, etc.). The model achieved satisfactory performance in classifying show vs. no-show patients.

Step 5 – Prediction & Dashboard

Predictions were exported as CSV and imported into Power BI to create an interactive dashboard showing patterns and key performance indicators (KPIs).

## 5. Conclusion

The project successfully identified significant patterns contributing to patient no-shows. A machine learning model was developed to assist healthcare providers in predicting and mitigating missed appointments. Power BI dashboards provide actionable insights to optimize scheduling strategies, such as sending timely reminders and prioritizing high-risk patients.

📌 Recommendations:

- Send SMS reminders for all patients.  
- Reduce waiting periods between scheduling and appointments.  
- Monitor trends by weekday and patient demographics.