Medical Appointment No-Show Prediction

1. Introduction

Missed medical appointments are a widespread challenge in healthcare, leading to operational inefficiencies, increased costs, and underutilized resources. This project leverages historical appointment data to build a machine learning model that predicts whether a patient is likely to miss their appointment. These insights aim to support smarter scheduling and improved patient outcomes.

2. Project Objective

The objective of this project is to develop a machine learning model that predicts whether a patient will miss their medical appointment. By analyzing historical data and identifying key behavioral patterns, the project aims to support healthcare providers in minimizing no-shows and improving scheduling efficiency.

3. Tools Used

- Python (Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib)
- Jupyter Notebook
- Power BI
- Microsoft Word (reporting)

4. Dataset Overview

The dataset includes over 100,000 medical appointment records from Brazil, with features such as patient demographics, health conditions (e.g., hypertension, diabetes), appointment scheduling details, and whether the patient showed up.

5. Methodology

The project followed an end-to-end data science workflow consisting of:

- Data Cleaning – Processed dates, encoded categories, removed invalid entries
- Exploratory Data Analysis (EDA) – Identified trends by age, SMS, weekday
- Feature Engineering – Created 'waiting_days' feature, encoded 'no_show'
- Model Training – Trained Decision Tree Classifier
- Model Evaluation – Measured using accuracy and F1-score
- Dashboard – Built Power BI visuals for real-time insights

6. Results & Insights

The Decision Tree model achieved approximately 78% accuracy. Factors like SMS reminders, patient age, appointment weekday, and waiting days significantly influenced no-show behavior. Power BI visualizations highlighted higher no-show rates on Mondays and among younger patients.

7. Recommendations

- Send SMS reminders at least 24–48 hours before appointments
- Use no-show history to flag high-risk patients for double confirmation
- Avoid scheduling on high-risk weekdays (e.g., Monday)
- Prioritize elderly patients with later appointment times

8. Conclusion

By integrating machine learning predictions and interactive Power BI dashboards, healthcare administrators can take proactive steps to reduce missed appointments. The project demonstrates a scalable approach for leveraging data to improve efficiency and patient care.