Healthcare Appointment No-show Prediction Project Report

1. Project Objective

Patient no-shows pose challenges to healthcare providers, causing wasted resources, increased costs, and reduced care efficiency. This project aims to predict patient no-shows using historical appointment data, uncover factors impacting attendance, and provide actionable insights to help reduce no-shows.

2. Dataset Description

The dataset contains over 110,000 appointment records from a healthcare provider, including:

- Patient demographics: Age, Gender, and Neighborhood.
- Health information: indicators for Hypertension, Diabetes, Alcoholism, and Handicap.
- Social support and communication: Scholarship status and whether an SMS reminder was sent.
- Appointment details: scheduled and actual appointment dates.
- Target variable: No-show status indicating whether the patient missed the appointment.

3. Data Cleaning and Preprocessing

- Removed duplicates and irrelevant columns.
- Converted date columns to datetime and extracted relevant features.
- Checked for missing or inconsistent values and handled them appropriately.
- Analyzed distributions and identified outliers in variables like Age.
- Encoded categorical variables to prepare for modeling.
- Balanced class distribution or noted imbalances for model choice considerations.

4. Exploratory Data Analysis (EDA)

- Visualized gender distribution, age groups, and neighborhood appointment volumes.
- Investigated chronic conditions and their relationship with no-shows.
- Explored impact of SMS reminders on attendance rates — found reminders correlate with higher attendance.
- Analyzed temporal patterns like appointment days and schedule lead times.
- Highlighted demographic trends affecting appointment adherence.

5. Model Development and Selection

- Selected a Decision Tree Classifier for interpretability and ease of explaining results.
- Split dataset into training and testing sets to validate generalization.
- Tuned model hyperparameters for balanced depth and accuracy.

- Trained the model on key predictors: Age, Scholarship, Hypertension, Diabetes, Alcoholism, Handicap, and SMS_received.

6. Feature Importance and Interpretation

- SMS_received emerged as the most impactful feature, confirming SMS reminders reduce no-shows effectively.
- Age is an important factor, with attendance patterns differing across age brackets.
- Other features including Scholarship, Handicap, Diabetes, Alcoholism, and Hypertension had moderate but meaningful influence.
- These insights help prioritize intervention strategies.

7. Model Evaluation and Predictions

- Evaluated the model using accuracy and confusion matrix metrics.
- Model demonstrated satisfactory predictive ability for no-shows, though some misclassifications remain.
- Sample prediction outputs show practical performance on unseen data.
- Provides a baseline for future enhancement with more complex or ensemble algorithms.

8. Business and Clinical Implications

- Healthcare providers can reduce no-shows by improving communication like SMS reminders.
- Targeted outreach and support for at-risk groups identified by chronic conditions and age can improve attendance.
- Predictive modeling supports optimized appointment scheduling, including strategic overbooking to accommodate likely no-shows.
- Improves resource utilization, staff planning, and patient access.