Healthcare Appointment No-Show Prediction — Project Report

Introduction

Missed appointments waste clinician time and delay patient care. Our project predicts which patients are likely to skip and visualises the patterns so a clinic can over-book intelligently or intervene with reminders.

Abstract

A synthetic dataset of 5 000 appointments (Jan–May 2025) was generated to mirror real-world behaviour: age spread, SMS reminders, weekday distribution and variable booking lead-times.
After cleaning and feature engineering, a Decision Tree (max_depth = 4) achieved **≈ 69 % accuracy** in classifying no-shows.
Key insights surfaced in Power BI reveal higher risk for children, seniors, weekend slots and bookings made > 30 days ahead.

Tools Used

- **Python**: Pandas, NumPy, scikit-learn (DecisionTreeClassifier)
- **Power BI Desktop**: interactive visuals & DAX measures
- **Joblib / CSV** for model & data portability

Steps Involved

1. **Data Generation & Cleaning**
    - Built 5 000-row CSV with Patient ID, Age, Gender, ScheduledDate, AppointmentDate, SMS_received, Weekday, NoShow.
    - Added engineered fields: Age Band, Lead Days, Lead Bucket.
2. **Modelling**
    - Split 80 / 20 stratified train-test.
    - Trained Decision Tree; pruned at depth 4 for interpretability.
    - Saved model as decision_tree_model.pkl.
3. **Power BI Dashboard**
    - Measures: **Total Appointments**, **Total No-Shows**, **No-Show Rate** plus segment rates.
    - Visuals: KPI card, No-Show Rate by Weekday column, Age Band × SMS stacked bar, Lead-Time combo, heat-map matrix, monthly trend.
    - Slicers for date, gender, age and SMS enable drill-downs.
4. **Insights & Recommendations**

- **SMS reminders** cut no-show odds by ~15 %.
- **Weekends** show 5–7 pp higher rates → slight over-booking advised.
- **Lead-time > 30 days** doubles risk → offer nearer slots.
- **Children (<12) & seniors (>60)** need follow-up calls.

Conclusion

Integrating predictive modelling with clear BI visuals equips staff to act before a slot is wasted. By adopting the highlighted interventions—more reminders, dynamic scheduling buffers and targeted outreach—a clinic can realistically lower no-shows, boost utilisation and improve patient outcomes.