**Healthcare Appointment No-Show Prediction**

**Introduction**

Patients missing scheduled healthcare appointments (so-called “no-shows”) are a common issue that wastes clinical resources and reduces care quality. No-shows lead to idle staff time, increased costs, and longer waitlists. Predictive analytics can help: by estimating each patient’s likelihood of missing an appointment, providers can implement targeted interventions (such as reminders or overbooking) to optimize scheduling and resource use. This project addresses the problem of appointment no-shows with the goal of forecasting missed visits and improving scheduling efficiency.

**Abstract**

The project uses a healthcare appointment dataset to predict which patients will miss their appointments. First, the data are imported and cleaned in Python (using pandas) to handle missing values and encode features. A decision tree classifier (from scikit-learn) is trained to distinguish show-ups from no-shows. The model is evaluated on test data and achieves modest accuracy, identifying **age** as a key predictor in this sample. In parallel, a Power BI dashboard is created to explore trends. The analysis focuses on factors such as whether the patient received an SMS reminder, patient age groups, and day-of-week patterns. For example, prior studies show that appointment reminders via text or phone can reduce no-shows. The combined approach (modeling plus visual trends) provides actionable insights for scheduling.

**Tools Used**

* **Python**: Pandas for data manipulation and preprocessing; scikit-learn for modeling (Decision Tree classifier).
* **Power BI**: For creating interactive dashboards to analyze appointment trends (no-show rates by day of week, by reminder status, etc.).
* **Decision Tree** model: Chosen for its simplicity and interpretability — decision trees generate clear, understandable rules for classification.

**Steps Involved in Building the Project**

1. **Data Preprocessing**: Load the appointment data from an Excel file. Check and clean the data by confirming there are no missing values. Convert date fields to datetime objects and derive features like the day of week and the interval between scheduling and appointment date.
2. **Feature Engineering**: Encode categorical variables (e.g. gender, SMS received) as numeric. Create new features as needed (for instance, ScheduledDayOfWeek and AppointmentDayOfWeek from date fields).
3. **Train/Test Split**: Split the data into training and test sets (e.g. 75/25 split) to evaluate generalization.
4. **Model Building**: Initialize and train a Decision Tree classifier on the training data. The decision tree is suitable for this classification task and provides interpretable decision rules.
5. **Model Evaluation**: Use the test set to evaluate accuracy and other metrics (e.g. confusion matrix, precision/recall). In our example, the model showed roughly 75% accuracy on a small test set, with **Age** emerging as the most important feature.
6. **Trend Analysis in Power BI**: Load the same data into Power BI and create visual reports. Analyze the no-show rate by factors such as **SMS reminder** (sent vs. not sent), **age group**, and **appointment day of week**. For instance, visualization might reveal higher no-show rates on certain weekdays or among younger patients.
7. **Optimization Recommendations**: Based on the model and dashboard insights, propose actions: e.g., send additional reminders to high-risk patients, schedule more flexible or overbooked slots when no-shows tend to peak. These data-driven suggestions aim to reduce missed appointments and make better use of clinic time.

**Conclusion**

This study demonstrates that data-driven prediction can help mitigate appointment no-shows. The Decision Tree model, while simple, achieved reasonable predictive accuracy and highlighted **age** as a key factor. The Power BI dashboard provided clear patterns, such as variations in no-show rates by weekday or reminder status, which can guide scheduling decisions. In line with prior research, using patient-specific predictions allows targeted interventions (like text reminders) and optimized scheduling to improve clinic efficiency. Overall, the model and analyses support better appointment management: by identifying likely no-shows, healthcare providers can adapt schedules (e.g. overbooking or time buffers) and outreach strategies to minimize wasted resources and improve patient care.

**Sources:** Project analysis and methodology are consistent with literature on appointment no-show prediction and intervention strategies