Healthcare Appointment No-Show Prediction

Introduction

Missed medical appointments result in inefficient use of healthcare resources and increased patient wait times. By leveraging data science techniques, this project aims to predict which patients are likely to miss appointments, allowing clinics to optimize scheduling and improve resource management.

Abstract

This project involves building a machine learning model to predict no-show behavior among patients based on historical appointment data. Using a decision tree classifier in Python, the model identifies key factors such as age, appointment day, and SMS reminders. Visual analysis using Power BI reveals actionable insights for improving attendance rates. Final recommendations target smarter appointment scheduling and patient engagement.

Tools Used

- Python: Scikit-learn (Decision Tree model), Pandas (Data wrangling)
- Power BI: Dashboard for visual trend analysis
- Jupyter Notebook: Model development and evaluation

Steps Involved in Building the Project

1. Data Collection & Cleaning
   - Loaded appointment data into Python.
   - Cleaned anomalies (e.g., negative ages, incorrect dates).
   - Removed redundant or irrelevant columns (e.g., patient ID).

2. Feature Engineering
   - Derived features like appointment weekday, days between booking and appointment.

- Encoded categorical variables (e.g., gender, SMS received).

3. Model Building

- Trained a Decision Tree classifier to predict "no-show" status.
- Split data into training and test sets for model evaluation.
- Metrics used: Accuracy, Precision, Recall.

4. Exploratory Analysis with Power BI

- Analyzed correlations between no-shows and features like age, neighborhood, SMS reminders.
- Identified high-risk groups (e.g., young adults, Monday appointments).

5. Optimization Recommendations

- Send SMS reminders closer to appointment date.
- Double-book high-risk slots to reduce resource underutilization.
- Implement follow-up confirmations for first-time visitors or elderly patients.

Conclusion

The model effectively predicts no-show probabilities using appointment and patient features. Power BI dashboards provided key insights into behavioral patterns. By integrating these predictions into scheduling systems, healthcare providers can reduce missed appointments, improving both efficiency and patient care.