Healthcare Appointment No-Show Prediction

Introduction

Missed appointments in healthcare facilities are a significant concern, leading to resource wastage, scheduling inefficiencies, and reduced quality of patient care. This project aims to predict patient no-shows using historical data and provide actionable insights to reduce their occurrence. By leveraging machine learning and data visualization, we aim to assist healthcare providers in improving scheduling efficiency and patient engagement.

Abstract

This project focuses on developing a predictive model to classify whether a patient will attend their scheduled appointment. Using a synthetic dataset simulating real-world healthcare appointments, the project involves data cleaning, feature engineering, and model training using a Decision Tree Classifier. The data includes key factors like patient age, gender, wait time between scheduling and appointment, and whether a reminder SMS was received. The analysis is complemented with a Power BI dashboard that provides visual insights and guides strategic decision-making for appointment optimization.

Tools Used

- **Python Libraries**:
    - Pandas, NumPy – Data preprocessing
    - Scikit-learn – Model training and evaluation
    - Matplotlib, Seaborn – Data visualization
- **Power BI**:
    - For interactive dashboards to display trends and no-show statistics

---

Steps Involved in Building the Project

1. **Data Simulation**
    - Created a synthetic dataset of 1000 patient appointments including features like appointment dates, SMS received, and no-show status.
2. **Data Cleaning**
    - Converted date formats, corrected inconsistent appointment times, and created derived features like WaitDays and DayOfWeek.
3. **Feature Engineering**
    - One-hot encoded categorical data such as day of the week and gender.
    - Calculated WaitDays as the gap between scheduling and appointment.

Steps Involved (Continued)

4. **Exploratory Data Analysis (EDA)**
    - Plotted age vs. no-show distribution
    - Analyzed effect of SMS reminders
    - Assessed wait times and their impact on attendance
    - Visualized weekday trends for appointment adherence
5. **Model Training**
    - Used a **Decision Tree Classifier** to predict no-shows.
    - Achieved a reasonable balance between precision and recall with a test accuracy of around 70-80%, depending on parameters.
6. **Power BI Dashboard**
   Exported the cleaned data to Power BI to build interactive visuals:
    - No-show rate by **weekday**, **age**, and **wait time**
    - Impact of **SMS reminders**
    - Trends in **no-show behavior** across demographic groups

---

Optimization Recommendations

- **Send timely SMS reminders**, especially for high-risk age groups or long wait times.
- **Avoid scheduling with long lead times** (reduce WaitDays).
- **Overbook slots** on days and times with historically high no-show rates.
- Focus on **weekday-specific strategies** to improve attendance (e.g., Mondays and Fridays tend to show more absences).

Conclusion

This project highlights how predictive modeling can support healthcare operational decisions. The Decision Tree model, combined with clear visual dashboards in Power BI, enables data-driven policies to reduce patient no-shows. Implementing the insights from this project can improve scheduling efficiency, reduce idle times for medical staff, and ultimately enhance patient care outcomes.