Predictive Analysis of Patient No-Shows in Healthcare Appointments

Done by: Sana Fathima

Introduction

Missed appointments burden healthcare systems, leading to financial losses and delayed treatments. This project utilizes historical appointment data to identify patterns that contribute to no-shows. Through predictive modeling and visualization, we aim to offer a data-driven solution to forecast patient attendance and optimize appointment scheduling.

Abstract

Patient no-shows lead to inefficient resource utilization and missed healthcare opportunities. This project aims to predict whether a patient is likely to miss a scheduled medical appointment using machine learning techniques and to uncover actionable insights through data visualization. The model helps healthcare providers improve scheduling efficiency and reduce missed appointments, ultimately enhancing patient care.

Tools Used

- **Python** (Pandas, scikit-learn, matplotlib, seaborn) for data preprocessing and model building
- **Power BI** for data visualization and trend analysis
- **Jupyter Notebook / VS Code** for code development and testing
- **MS Excel / CSV** format for data handling

Steps Involved

1. **Data Collection & Cleaning**
    - Imported the dataset (appointments.csv)
    - Removed nulls, corrected data types, and created new features (e.g., days_waiting)
2. **Feature Engineering**
    - Encoded categorical variables (like gender, appointment_dayofweek)
    - Created a binary label (no_show) and calculated waiting time between scheduling and appointment
3. **Model Development**
    - Trained **Logistic Regression** and **Random Forest** classifiers
    - Evaluated model using **accuracy, precision, recall, and F1-score**

- Logistic Regression achieved the most balanced results (Accuracy: ~66.8%) but since we were more focused on NO-Show so we choose to go with **Random Forest** Model

4. **Insight Generation in Power BI**
    - Built dashboards to explore trends across age, SMS reminders, scholarship status, chronic diseases, waiting time, and weekdays
    - Used slicers and charts for interactive insights
5. **Recommendations**
    - Target patients in neighbourhood with high no-show rate.
    - Use SMS reminders more effectively for young adults and scholarship beneficiaries.
    - After a missed appointment, send a quick feedback form: Was the timing wrong? Reminder missed? Transportation issue?
    - For patients with chronic disease offer flexible re-scheduling.
    - Week days like Monday, Tuesday and Wednesday have high no-show rate probably due to transportation issues or missed appointments due to workload.
    - Reduce appointment density on those days and over-book slightly on other days for routine follow-ups.

Conclusion

The project successfully developed a prediction model to identify potential no-shows in healthcare appointments. Through data analysis and visualization, critical insights were obtained to help healthcare providers mitigate no-shows and enhance operational efficiency. Combining predictive analytics with targeted scheduling interventions can drive better patient outcomes and resource planning.