Healthcare Appointment No-Show Prediction

Name: Rama Subrahmanyam Vudatha

Introduction

In healthcare systems, missed appointments lead to inefficiencies, resource wastage, and delayed care. Predicting whether a patient will attend their appointment helps hospitals plan better and reduce no-show rates. This project aims to build a predictive model to identify patients who are likely to miss their appointments using machine learning.

Abstract

This project analyzes a real-world dataset of over 100,000 medical appointments in Brazil. Each record includes demographic details, appointment information, and whether the patient showed up or not. Using this data, we trained a Decision Tree Classifier to predict no-shows. Additionally, insights were visualized using Power BI to support better scheduling decisions.

Tools Used

- Python: Pandas, Seaborn, Sklearn (for modeling and EDA)
- Power BI: For interactive dashboards and visual insights
- Jupyter Notebook: For model building and evaluation

Steps Involved in Building the Project

1. Data Collection: Dataset sourced from Kaggle ("No-show appointments").
2. Data Preprocessing:
   - Removed irrelevant columns like AppointmentID and ScheduledDay
   - Converted categorical data (Gender, No-show) into numerical format
   - Removed invalid age values (negative values)
3. Exploratory Data Analysis (EDA):
   - Found that factors like age, SMS reminders, and health conditions affect show rates.
   - Visualized trends using bar graphs, pie charts, and heatmaps.
4. Model Building:
   - Trained a Decision Tree Classifier with 79.9% accuracy
   - Evaluated using confusion matrix, classification report, and accuracy score
5. Visualization with Power BI:

- Built dashboards showing no-show rates by gender, age, SMS reminder, etc.
- Included filters and slicers for better insight

Conclusion

The model showed that patients who did not receive SMS reminders, were young, or had no chronic conditions were more likely to miss appointments. This suggests that:
- Sending reminders increases show-up rate
- Prioritizing patients by no-show risk improves scheduling
- Resource allocation can be optimized with predictive insights

This project proves how data analytics can make healthcare more efficient and proactive.