Project Report: Healthcare Appointment No-Show Prediction

Introduction:

Patient no-shows at their scheduled appointments pose a significant challenge in healthcare organizations, leading to inefficient resource allocation and impacting patient access to care. This project aims to analyse historical appointment data to identify key factors influencing whether a patient attends their scheduled appointment. By understanding these drivers and predicting no-show likelihood, we seek to optimize scheduling practices to improve overall attendance.

Abstract:

This project employs Python, Power BI, and machine learning to analyse a dataset of patient appointments to identify critical factors associated with missed appointments. Utilizing data analysis and predictive modeling, we found that scheduling lead time, patient age, and neighbourhood are key determinants of no-show behavior. The insights gained will inform strategies to predict high-risk appointments and optimize scheduling for better resource utilization and reduced no-shows.

Tools Used:

- Python (Pandas, Matplotlib, Seaborn, Sklearn, Numpy) - Data preprocessing, model building, feature importance analysis.
- Power BI - designing interactive dashboard for attrition visualization.
- Decision Tree Analysis – Identifying the most important factors influencing a patient's likelihood of not showing up.

Steps Involved:

Step 1: Data Exploration (EDA) in Python:

- Analysed Neighbourhood-wise no-show trends using bar charts.
- Analysed Gender-wise no-show trends using bar charts.
- Analysed Age group-wise no-show trends using bar charts.
- Analysed Data-difference wise no-show trends using bar charts.
- Analysed SMS Alert receipt-wise no-show trends using bar charts.

Step 2: Built Decision Tree Model:

- Split data into training/testing sets.
- Trained Decision Tree model to predict no show.
- Examined model performance and feature importance of each column.

Step 3: Visualized No-show Factors Using Power BI:

- Used Bar Chart for Top 10 Neighbourhoods with no-show patients.
- Used Column Chart for Weekday-wise patient no-show counts.
- Used Pie chart to display SMS Alert receipt wise patient no-show counts.
- Used Donut chart to display Age group wise patient no-show counts.
- Used Column Chart to track Date difference wise patient no-show counts.

Final Takeaways:

- The dataset highlights a considerable challenge with healthcare appointment attendance, showing an overall no-show rate of approximately 20%. This indicates a widespread issue that could impact resource allocation and patient care access.
- While visuals might show differences in no-show counts based on factors like gender or SMS receipt, the predictive model indicates that scheduling time, age, and neighbourhood are considerably more influential in determining whether a patient will show up compared to gender, medical conditions, or the status of receiving a single SMS alert.
- The time elapsed between scheduling an appointment and the actual appointment day is a critical factor in predicting no-shows. Appointments scheduled on the same day or just one day in advance have the highest absolute numbers of no-shows, and 'Date.diff' is identified as the most important feature in the predictive model.
- Younger age groups (especially below 30) account for the largest counts of no-shows, and specific neighbourhoods show disproportionately high numbers of missed appointments, making these crucial factors for targeted interventions.