PROJECT REPORT ON

**HEALTHCARE APPOINTMENT NO-SHOW PREDICTION**

**INTRODUCTION**

A "medical no-show" refers to a patient who fails to attend a scheduled medical appointment without providing prior notice. This can lead to wasted resources, decreased efficiency, and potential harm to the patient's health due to disrupted treatment plans. A no-show is distinct from a cancellation. A cancellation is when a patient informs the practice, they will not be able to attend their appointment in advance.

**ABSTRACT**

This study presents an exploratory analysis of the dataset containing various variables related to patient health and diagnostic information. It includes variables like age, hypertension status, diabetes, alcoholism and other clinical indicators relevant to disease detection and risk assessment. The objective is to predict whether patients are likely to miss their appointments thereby enabling efficient scheduling of the appointment. The findings may contribute to optimizing appointment systems by providing data-driven recommendations for targeted reminders and improved patient engagement strategies.

**TOOLS USED**

Python- Jupyter notebook (Libraries used: Sklearn, Pandas), MS Powe BI, MS Word, MS Excel

**STEPS INVOLVED**

1. **Data exploration and data cleaning:**

Imported necessary libraries then imported data, checked data type and null values. Encoded the variables Gender, No-show, Neighbourhood using label encoder.

1. **Summary statistics:**

* The scheduled day and appointment columns show that most of the appointments were scheduled just a few days before actual appointment date and some on the same day.
* Most patients do not suffer from chronic illness and are not receiving scholarships.
* Some patients who did not receive a SMS reminder regarding the appointment still attended their appointments while some missed it even after receiving a reminder through SMS.
* Overall, we observe most patients attended their appointments.

1. **Heatmap and boxplot conclusions:**

* We observe that most features have very weak correlations with no-show variable. Positive correlation is with SMS\_recieved (0.13) which suggests that receiving a SMS may slightly increase the likelihood of missing an appointment possibly due to targeting high-risk patients. Other variables such as age, gender, and health conditions (e.g., hypertension, diabetes) show negligible influence on no-show behaviour, indicating that no single factor stronglypredicts absenteeism.
* The age distribution of patients is fairly wide, ranging from near 0 to over 100. The median age is around 35-40 years and most patient lie within 18 to 55 age group. A few outliers exist above 100 suggesting possible data entry errors or rare valid case.

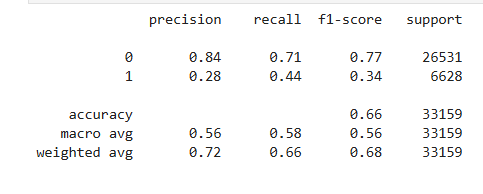
1. **Distribution plot/ histogram conclusions:**

The distribution plot indicate that the dataset is heavily imbalanced across several binary variables. Most patients do not have conditions like hypertension, diabetes, alcoholism, or disabilities (Handicap) and only a minority received scholarships or SMS reminders.

1. **Decision Tree Classifier:**

After splitting the data into training and testing data we fit the decision tree classifier model for classification.

Classification report:



**Conclusion:**

* We observe significant class imbalance and performance disparity between the two classes (0 and 1).
* An overall accuracy of 66% was achieved.
* Model performs well on Class 0 with 0.84 precision and 0.77 F1-score.
* Model struggles to correctly identify class 1 with poor precision 0.28 and F1-score 0.34.
* The macro average has a F-1 score of 0.56 highlighting poor overall balance.
* The weighted average has a F1- score of 0.68 indicating model is skewed toward class 0 performance due to its larger share.

**DASHBOARD INSIGHTS:**

We observe that majority of patients attended their appointments (approx. 80%) with 20% no shows. Female patients represent a larger portion of the dataset. We observe majority of patients with no chronic illness. Most of them did not receive SMS reminders yet attended the appointment. Also, majority of patients do not receive scholarships. SMS reminders, weekday scheduling or no prior (at least a week before) SMS regarding the appointment may influence no-shows. Most no-shows occur among non-scholarship, healthy and non-reminded patients.

**RECOMMENDATION:**

* Target SMS or call reminders based on model predictions.
* Penalty or incentives policies for no-shows or no-prior notice regarding absence.
* Inform patients about their next appointment date/time before they leave.
* Allow patients to reschedule or cancel easily via phone, app, or website.