Project Title:Healthcare Appointment No-Show Prediction

* Introduction:

This project aims to analyze and reduce the no-show rate in healthcare appointments. By studying historical appointment data, it focuses on identifying the key factors that influence whether patients attend their scheduled visits. The goal is to uncover trends related to age, appointment day, and communication methods like SMS reminders. Insights from this analysis can help healthcare providers optimize scheduling and improve overall patient engagement.

* Abstract:

Missed appointments in the healthcare sector lead to inefficient use of resources, increased patient wait times, and reduced quality of care. This project focuses on predicting patient no-shows using historical appointment data. By analyzing variables such as age, appointment day, and SMS reminders, the project identifies patterns that contribute to appointment non-attendance. A decision tree model is developed to predict the likelihood of no-shows, and high-risk cases are flagged for potential intervention. The insights generated aim to support healthcare providers in optimizing appointment scheduling and improving patient compliance

* Tools Used:

1)Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn):

For data cleaning, preprocessing, exploratory data analysis, model building, and evaluation.

2)Google Colab:

Used as the coding environment to implement and test the machine learning model.

3)Power BI:

For creating interactive dashboards and visualizations to showcase insights and prediction results.

4)Microsoft Excel:

Used for initial data inspection, editing, and exporting cleaned or filtered datasets.

* Steps Involved in Building the Project

🛠 Data Cleaning & Preparation

-Removed duplicates & invalid ages.

-Converted date columns to datetime format.

-Renamed inconsistent columns.

-Verified no missing values.

🔍 Feature Engineering

Derived: AppointmentWeekday, ScheduledWeekday, DaysBetween.

Encoded:

-No\_Show: 1 = No-show, 0 = Show

-Gender, Neighbourhood: One-hot encoding

🧪 Model Training

✅ Decision Tree Classifier

Params: class\_weight='balanced', max\_depth=6, min\_samples\_split=10

-Accuracy: 58%

-Recall (No-show): 81% (high), but low Precision: 30%

Useful for identifying no-shows, but overpredicts them.

🔁 Random Forest Classifier

Params: n\_estimators=100, class\_weight='balanced'

-Accuracy: 81%

-Recall (No-show): 14%, Precision: 58%

-Better balance; fewer false alarms, more stable.

📊 Insights

-SMS Reminder: Higher no-show (27.6%) with SMS → Reminders alone ineffective.

-Age: 19–30 age group had highest no-show (24.6%); older patients more reliable.

-Weekday: More no-shows on Saturday, fewer midweek → Schedule more midweek.

🚀 Real-Time Use

-Input new patient data → Predict no-show risk (e.g., 61.67% chance of no-show).

-Enables preventive steps like early rescheduling or targeted reminders.

📁 Output

-Final predictions and insights saved as CSV for further use.

* Conclusion:

This project successfully utilized machine learning to predict patient no-shows, identifying key factors like age, SMS reminders, and appointment day. By comparing Decision Tree and Random Forest models, it was found that Random Forest provided better accuracy and fewer false positives, though challenges remain in accurately predicting no-shows. Insights suggest that SMS reminders alone are ineffective, younger patients are more prone to no-shows, and scheduling midweek appointments may improve attendance. The model can now be used for real-time predictions, allowing healthcare providers to take proactive steps in optimizing patient attendance and reducing wasted resources.