INTRODUCTION

Missed medical appointments create inefficiencies in healthcare systems, wasting time and resources while affecting patient care. This project aims to develop a predictive model to identify patients likely to miss their appointments. Such predictions enable healthcare providers to optimize scheduling and improve attendance rates.

ABSTRACT

This project leverages machine learning and data visualization to address the challenge of appointment no-shows. A decision tree classifier is trained on patient appointment data to predict no-shows based on features such as SMS reminders, age, and scheduled weekday. Insights gained are visualized in a Power BI dashboard to support data-driven decisions. Optimization recommendations are provided to improve scheduling and patient communication strategies.

TOOLS USED

Python: Data preprocessing, model training, evaluation

Libraries: Pandas, Scikit-learn, Matplotlib

Power BI: Visualization and trend

STEPS INVOLVED IN BUILDING THE PROJECT:

Data Collection and Import

Appointment data was imported in CSV format and read using Pandas.

2. Data Cleaning and Preprocessing

Removed duplicate and irrelevant columns

Converted date columns to datetime format

Handled missing and inconsistent data

Created binary target column: No-show (Yes/No → 1/0)

3. Feature Engineering

Extracted features: age, gender, scheduled day, appointment day, day of the week, SMS received, scholarship

Removed outliers (e.g., negative ages)

4. Model Building

Split data into training and test sets

Trained a Decision Tree Classifier

CONCLUSION

This project utilized Python and Power BI to address the problem of missed healthcare appointments. Using Python, we built a decision tree model that predicted patient no-shows based on features such as age, SMS reminders, and appointment scheduling details. The results were further explored through interactive visualizations in Power BI, revealing key trends and patterns. These insights can help healthcare providers improve scheduling efficiency by targeting patients more likely to miss appointments and implementing timely reminder strategies. Overall, the project demonstrates how combining Python-based modeling with Power BI visualization can support data-driven decision-making in healthcare. Libraries: Pandas, Scikit-learn, Matplotlib. Power BI: Visualization and trend analysis