

HEALTHCARE APPOINTMENT NO- SHOW PREDICIONS

Introduction:

In the healthcare sector, missed appointments or no-shows pose serious problems since they result in resource waste, increased wait times, and less patient access. With the help of data analytics and machine learning, this initiative seeks to forecast patient attendance at appointments and suggest scheduling adjustments that increase productivity. Data-driven decision-making and clinical management are supported by the combination of Power BI for visualization and Python (Scikit-Learn, Pandas) for model construction.

Abstract:

To estimate the probability that a patient will miss an appointment, the study examines medical appointment data. The model finds patterns impacted by variables including age, lead days, appointment day, SMS reminders, and patient health issues using a Decision Tree Classifier. To give healthcare management useful information, the generated predictions are shown in an interactive Power BI dashboard.

Interventions like automated SMS reminders or rescheduling can be used to target predicted high-risk patients, lowering total no-show rates and improving operational effectiveness.

Tools Used:

Data Handling & Cleaning	Python, Pandas, NumPy
Modeling & Prediction	Scikit-Learn (DecisionTreeClassifier)
Visualization & BI	Power BI Desktop
Data Presentation	Matplotlib, Seaborn, Power BI charts
Documentation	Jupyter Notebook, Microsoft Word / PDF

Steps Involved in Building the Project:

a. Gathering and loading data

The dataset was gathered from publicly available sources, including the Medical Appointment No-Show dataset on Kaggle.

Age, gender, scheduled day, appointment day, and SMS reminders were among the demographic and behavioral characteristics included in the data.

b. Data Preparation and Cleaning

Null values and duplicates were eliminated.

Date columns were converted to DateTime format.

DaysInAdvance, AppointmentWeekday, and TotalConditions are examples of newly added engineering features.

Categorical variables were encoded for use in the model.

c. Using Seaborn and Matplotlib, exploratory data analysis (EDA)

visualized no-show trends by age group, weekday, and SMS reminders.

Shorter scheduling windows, appointment days, and reminders were found to significantly lower the likelihood of no-shows.

d. Model Training and Prediction

Using Scikit-Learn's Decision Tree Classifier:

NoShow_Binary is the target variable (1 = missed, 0 = attended).

Model accuracy was between 80 and 85 percent.

Generated anticipated probabilities (NoShowProbability) for each patient.

e. Results (appointment_predictions.csv) were imported into Power BI.

made a number of interactive pages: KPIs for the total number of appointments, model accuracy, and current no-show rate are displayed on the summary page.

Risk Analysis Page: Displays high-risk patients using a slicer by Risk Level.

Factor Impact Page: Displays daily trends and the SMS effect.

The optimization page uses lead-time visualizations and target gauge charts to suggest scheduling efficiency.

f. Suggestion and Enhancement

The target decreased the no-show rate by 15%.

Use short advance scheduling (less than 14 days), overbooking (1 per 6 high-risk patients), and targeted SMS notifications. track continuous performance in relation to this goal, use Power BI reports.

Conclusion:

The Healthcare Appointment No-Show Prediction Project is an effective example of how data science may be used to solve operational issues in the medical field.

By anticipating no-shows, medical facilities can:

Cut down on 15–25% of missed appointments.

Improve patient communication by taking preventative measures.

Increase scheduling effectiveness and make the best use of available resources.

This project gives healthcare organizations a dynamic visualization framework and a predictive model so they can act quickly on findings.