

Elevate Labs

Internship

Final Project - Report

Healthcare Appointment No-Show Prediction

Sireesha Karagana

(MCA in AI & ML)

18 May 2025

Predicting No-Shows for Medical Appointments Using Machine Learning

1. Introduction

Missed medical appointments create inefficiencies in healthcare, leading to wasted resources, poor patient care, and increased operational costs. Predicting these no-shows in advance helps hospitals optimize scheduling and improve service delivery. This project uses machine learning models to predict whether a patient will attend a scheduled medical appointment, based on demographic, clinical, and appointment-related features.

2. Abstract

This project analyzes a dataset of 110,527 medical appointments in Brazil to identify patterns associated with patients who miss their appointments. Using Python and machine learning techniques, the data was preprocessed and cleaned. Exploratory Data Analysis (EDA) revealed important trends regarding age, gender, medical history, SMS reminders, and other factors. Predictive models like Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN) were trained and evaluated. Among them, the Random Forest Classifier achieved the best accuracy of 79%. This work highlights the potential of machine learning in supporting healthcare scheduling systems.

3. Tools Used

Programming Language: Python

Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

Development Environment: Jupyter Notebook

4. Steps Involved in Building the Project

◆ Data Collection:

The dataset was downloaded from Kaggle and includes features like Patient ID, Age, Gender, Scheduled Day, Appointment Day, SMS_received, and No-show status.

◆ Data Cleaning & Preprocessing:

Converted date columns to datetime objects.

Handled inconsistent or incorrect data (e.g., negative ages).

Encoded the target variable 'No-show' as binary.

Removed irrelevant columns like 'PatientId' and 'AppointmentID'.

◆ Exploratory Data Analysis (EDA):

Analyzed age groups and their no-show behavior.

Compared attendance across gender, SMS reminders, and scholarship status

Visualized no-show distribution using bar plots and count plots.

Observed trends showing younger patients and those without SMS reminders were more likely to miss appointments.

◆ Model Building:

Three classification models were trained:

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Random Forest Classifier

◆ Model Evaluation:

- Split the data into training and testing sets (80-20).
- Used accuracy, precision, recall, and confusion matrix to evaluate models. Random Forest performed best with ~79% accuracy and balanced precision/recall.

5. Conclusion

The project successfully demonstrated how machine learning can predict no-shows in medical appointments. Random Forest emerged as the best-performing model. Important features influencing attendance included age, SMS reminders, and medical conditions. Future improvements could involve incorporating time-of-day and weather data. Such predictive systems can help clinics reduce missed appointments and improve efficiency.