

Healthcare Appointment No-Show Prediction

Project Report

Introduction: This project utilises healthcare appointment data to analyse the frequency of no-shows despite prior appointments. The model predicts whether patients will miss their scheduled appointments despite SMS notifications and generates actionable recommendations to improve scheduling and reduce no-shows.

This project uses the "Medical Appointment No-Shows" dataset from Kaggle, containing 110,527 appointment records from public hospitals in Brazil. Key variables include patient demographics, appointment dates, and whether the appointment was attended.

Tools and Technologies:

1. **Python:** Pandas, NumPy, Matplotlib, Scikit Learn (Decision Tree Classifier and Random Forest Classifier, Confusion Matrix), Seaborn.
2. **Power BI:** Bar Chart, Donut Chart, Stacked Column Chart, Slicers.

Steps Involved in the Project:

Data Cleaning:

- Renamed columns for consistency
- Converted dates to datetime format
- Created new features: weekday of appointment, waiting days
- Encoded categorical variables (e.g., gender, no-show)
- Removed records with negative waiting times

Modelling:

- A Decision Tree Classifier is used to train the model to predict no-shows using features such as age, gender, waiting days, and SMS received. A confusion matrix is created using the testing and predicted values.
- The confusion matrix predicted 15424 true negatives, ie, showed prediction correctly. The model showed no false positives and no true positives. The model showed 6164 false negatives as evident from the confusion matrix.
- The classification report is prepared using a Random Forest Classifier from Scikit Learn. The report reflects that 73% of patients showed up and of

those 70% were correctly identified. Similarly, 32% of patients did not show up (high false positives) and of them, 36% were caught (high false negatives). The overall accuracy of the model is 60%.

- A feature importance graphic is also created in which it is surfaced that age and waiting days are the most important features affecting the turnout of patients.

Exploratory Data Analysis:

- The age-based trends show that the highest no. of no-shows is among the adults.
- The stacked column chart reflects that Wednesday has the highest no. of no-shows.
- The SMS notifications bar chart shows there is no significant effect of SMS alert on patient turnout.
- The donut chart shows there are higher chances of a female patient of not showing up.

Optimised Recommendations:

- Send SMS reminders approximately 48 hours before and then again 24 hours before the appointments.
- Avoid booking new patients on high no-show days, such as Wednesday and Monday.
- Give priority slots to seniors and chronic patients on Wednesdays to reduce the higher no-show ratio.
- Minimise waiting time between scheduling and appointment.
- Consider overbooking on high-risk no-show days such as Wednesdays.

This project reflects the use of machine learning and data visualisation to improve and support effective healthcare scheduling. The predictive model offers strong interpretability, and the Power BI dashboard enables data-driven decision-making to increase patient turnout ratio.