

# Healthcare No-Show Prediction - Project Report

Healthcare No-Show Prediction

-----

Project Type: Machine Learning

Domain: Healthcare

Language: Python

Tools Used: Pandas, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebook

## Project Overview:

Missed medical appointments, or "no-shows," cause scheduling inefficiencies, increased healthcare costs, and wasted medical resources.

This project aims to predict whether a patient will show up for their scheduled appointment using historical appointment data and machine learning models.

## Dataset Summary:

The dataset was sourced from Kaggle and contains 110,000+ medical appointment records from Brazil.

## Key Features:

- Gender
- ScheduledDay and AppointmentDay
- Age
- Neighbourhood
- Scholarship (social welfare indicator)
- SMS\_received
- No-show (target variable)

# Healthcare No-Show Prediction - Project Report

## Data Preprocessing Steps:

- Converted date columns to datetime format and calculated waiting time.
- Removed outliers and invalid data (e.g., negative ages).
- Encoded categorical variables (e.g., Gender, No-show).
- Handled class imbalance using SMOTE and class weights.
- Normalized numeric features to improve model performance.

## Model Building:

Several machine learning models were tested:

Model	Accuracy	Precision	Recall	F1-score
-----				
Logistic Regression	79%	72%	63%	67%
Random Forest	84%	76%	74%	75%
XGBoost	85%	78%	76%	77%

XGBoost performed best and was selected for final deployment.

## Key Insights:

- Patients with longer waiting times are more likely to skip appointments.
- Receiving an SMS reminder slightly reduces the chance of a no-show.
- Younger patients and those without health coverage are more likely to miss appointments.
- Certain neighborhoods show higher no-show rates, indicating location-based trends.

## Optimization Recommendations:

## Healthcare No-Show Prediction - Project Report

1. Feature Engineering: Add features like appointment type, distance from clinic, or previous no-show history.
2. Model Improvements: Apply hyperparameter tuning using GridSearchCV and ensemble methods.
3. Interpretability: Use SHAP values to explain individual predictions to stakeholders.
4. Operational Integration: Deploy the model in hospital systems to flag high-risk patients and automate reminders.

### Conclusion:

This project demonstrates how machine learning can be used to reduce no-show rates and enhance hospital efficiency.

By accurately predicting patient attendance, healthcare providers can improve scheduling and reduce costs.

### Future Work:

- Deploy the model using Flask or Streamlit for a live prediction tool.
- Connect with a real-time database to update patient information.
- Integrate model with SMS/email APIs for alerting patients at risk of no-shows.