Project Report

Healthcare Appointment No-Show Prediction Using Machine Learning and Power BI

1. Introduction

In many healthcare systems, missed appointments (or “no-shows”) create significant problems — they waste valuable clinical resources, reduce operational efficiency, and can even affect patient health outcomes. This project aims to analyze a real-world dataset of medical appointments and predict whether a patient will attend their appointment. The final outcome includes both a predictive model using Python and a business intelligence dashboard using Power BI to assist in decision-making.

2. Objective

The primary objectives of this project are:

- To develop a machine learning model to **predict the likelihood of a patient missing their appointment**.
- To **analyze patterns and trends** that influence no-shows.
- To **visualize key insights** using Power BI for better understanding and stakeholder communication.
- To **recommend strategies** to reduce appointment no-shows.

3. Tools and Technologies

Tool/Library	Purpose
Python (pandas, matplotlib, seaborn, sklearn)	Data analysis, modeling, and visualization
Power BI	ractive dashboard creation
Kaggle Dataset	Source of real-world appointment data

4. Dataset Overview

- **Source:** Kaggle - Medical Appointment No Shows Dataset
- **Size:** ~110,000 medical appointment records
- **Location:** Brazil
- **Important Columns:**
  - PatientId, AppointmentID
  - ScheduledDay, AppointmentDay
  - Age, Gender, Neighbourhood
  - Scholarship, Hypertension, Diabetes, Alcoholism, Handicap
  - SMS_received
  - No-show (target variable)

5. Data Cleaning and Preprocessing

Performed multiple cleaning steps to prepare the data for analysis:

- **Removed irrelevant columns:** PatientId, AppointmentID

- **Date Conversion:** Converted ScheduledDay and AppointmentDay to datetime objects
- **Feature Created:**
    - WaitingTime = AppointmentDay - ScheduledDay (number of days between scheduling and appointment)
    - DayOfWeek = AppointmentDay.day_name() (weekday name of the appointment)
- **Handled inconsistencies:**
    - Removed rows with negative age values
- **Target Encoding:** Mapped the No-show column:
    - "Yes" → 1 (No-show)
    - "No" → 0 (Show)

6. Exploratory Data Analysis (EDA)

Analyzed key variables using Python visualizations:

- Count of no-shows by day of the week
- Distribution of Age vs No-show
- SMS reminders vs No-show rates
- Waiting Time and its effect on attendance

7. Feature Engineering and Model Building

Selected Features:

- Age
- SMS_received
- WaitingTime

Model Used:

- **Decision Tree Classifier** from scikit-learn

Train-Test Split:

- 70% training, 30% testing

8. Power BI Dashboard

Exported the cleaned data using:

python

Copy code

df.to_csv("cleaned_appointment_data.csv", index=False)

Visualizations in Power BI:

- Bar Chart: No-show count by weekday
- Donut Chart: SMS_received vs No-show
- Scatter Plot: Age vs Waiting Time (colored by No-show)
- Slicers: Age Range, Gender, Neighbourhood, Chronic Conditions

Dashboard Benefits:

- Easy to identify days with high no-show rates
- Understand how SMS and waiting time impact attendance
- Filter data by region, age group, or chronic condition

---

9. Key Findings

Factor	Observation
Waiting Time	Longer waiting time correlates with higher no-shows
SMS Reminder	Patients who received SMS were more likely to attend
Day of the Week	Mondays and Fridays had higher no-show rates
Age	Younger patients had more no-shows than older ones

---

10. Optimization Recommendations

1. Send multiple SMS or phone reminders, especially to younger patients.
2. Schedule appointments within shorter waiting periods to reduce forgetfulness or lack of motivation.
3. Avoid critical appointments on high no-show days (like Mondays).
4. Provide reminders in local languages or via WhatsApp for improved effectiveness.

---

11. Conclusion

This project demonstrated the power of data analytics and machine learning in solving real-world problems in healthcare. By identifying no-show patterns, we can recommend actionable

steps to improve appointment attendance, thus enhancing hospital efficiency and patient care.

The model provided interpretable predictions, and Power BI dashboards offered meaningful insights for healthcare administrators.

---

12. Future Enhancements

- Use advanced models like Random Forest, XGBoost, or ensemble learning
- Deploy as a web application with real-time predictions
- Include external factors like weather, traffic, and public holidays
- Integrate patient feedback or communication preferences

---

13. Deliverables

- Cleaned and preprocessed dataset
- Trained Decision Tree model
- Power BI interactive dashboard
- Optimization recommendation report

Report by

KARTHEEKHA C.V