Healthcare Appointment No-Show Prediction and Analysis

1. Introduction

Missed medical appointments are a major challenge for healthcare organizations, leading to wasted doctor time, increased operational costs, and reduced patient care efficiency.
This project focuses on analyzing patient appointment data to understand no-show behavior and predict the likelihood of a patient missing an appointment using machine learning techniques.

In addition, an interactive Power BI dashboard is developed to visualize trends and support data-driven decision-making.

2. Problem Statement

Hospitals face frequent appointment no-shows, which negatively impact resource utilization and patient service quality.
The objective of this project is to:

- Identify factors influencing appointment no-shows

- Predict patient no-show probability

- Provide actionable insights to reduce missed appointments

3. Dataset Description

The dataset contains historical healthcare appointment records with the following attributes:

- Age

- Gender

- SMS Reminder Received

- Weekday of Appointment

- No-Show Status (0 = Attended, 1 = No-Show)

This dataset is used for both predictive modeling and trend analysis.

4. Data Preprocessing and Feature Engineering

The following preprocessing steps were applied:

- Removal of missing and inconsistent data

- Encoding of categorical variables

- Creation of derived features such as:

    o Age Groups

    o Weekday

- Conversion of target variable (No-Show) into numerical format

These steps ensured the dataset was suitable for machine learning and visualization.

5. Exploratory Data Analysis and Trend Analysis

Trend analysis was performed to understand patient behavior:

SMS Reminder Analysis

Patients who received SMS reminders had a lower no-show rate, indicating the effectiveness of reminder systems.

Age-Based Analysis

Certain age groups showed higher no-show tendencies, highlighting the need for age-specific engagement strategies.

Weekday Analysis

No-show rates varied by weekday, suggesting that appointment timing influences patient attendance.

6. Machine Learning Model Development

A supervised machine learning approach was used to predict no-show behavior.

- Algorithms used:
    - Decision Tree
    - Random Forest
- Model evaluation metrics:
    - Accuracy
    - Confusion Matrix

The trained model helps identify high-risk appointments in advance.

7. Power BI Dashboard Design

An interactive Power BI dashboard was developed with:

- KPI Cards:
    - Total Appointments
    - Total No-Shows
    - No-Show Percentage
- Visualizations:
    - No-Shows by Age Group
    - No-Shows by Gender
    - No-Shows by Weekday
- Interactive Slicers:

- o Gender

- o SMS Reminder

- o Weekday / Age

The dashboard enables users to quickly analyze patterns and take informed decisions.

8. Optimization Recommendations

Based on insights from data analysis and dashboard findings, several optimization strategies were proposed, including improved SMS reminder strategies, weekday-based scheduling, age-based engagement, controlled overbooking, and appointment confirmation systems.

9. Conclusion

This project demonstrates how data analytics and machine learning can be effectively applied to address healthcare operational challenges.
By combining predictive modeling with interactive visualization, hospitals can reduce appointment no-shows, improve scheduling efficiency, and enhance patient care.

10. Future Scope

- Integration of real-time appointment data

- Use of advanced models such as Gradient Boosting

- Inclusion of additional patient behavior features

- Automation of alert systems for high-risk appointments