Healthcare No-Show Appointment Prediction Project Report

Abstract

Missed medical appointments are a persistent problem in healthcare systems worldwide, contributing to inefficiencies, increased operational costs, and compromised patient care.
To address this issue, this project aims to predict whether a patient will attend their scheduled medical appointment using machine learning. The approach combines data preprocessing, feature engineering, and a decision tree classifier.
Power BI is utilized to visualize the data and provide insightful dashboards that help understand the behavioral patterns of patients.
The goal is to support healthcare providers in proactive scheduling and patient engagement strategies.

Introduction

The increasing rate of missed medical appointments presents a significant challenge for healthcare professionals.
Understanding the factors behind patient no-shows can lead to better scheduling, reduced wait times, and improved resource utilization.
This project explores a dataset containing over 100,000 appointment records from Brazil to build a predictive model.
The main objective is to identify patients at high risk of missing their appointments, thereby allowing clinics to take preventive actions such as reminders or rescheduling.
This study leverages data science and visualization tools to derive actionable insights.

Tools & Technologies Used

 Python: Used for data loading, preprocessing, feature engineering, and building the machine learning model.

 Pandas & NumPy: For data manipulation and numerical operations.

 Matplotlib & Seaborn: Used to visualize data distributions and relationships.

 Scikit-learn: To train and evaluate the Decision Tree classifier.

 Power BI: For creating interactive dashboards to explore the impact of demographics and scheduling patterns on no-show rates.

 Jupyter Notebook: Environment used for experimentation and prototyping of the solution.

Steps Involved

1. Data Loading: Dataset from Kaggle loaded and explored.

2. Data Cleaning: Converted date columns, removed invalid ages, and mapped no-show labels.

3. Feature Engineering: Created features like weekday, waiting days; encoded categorical data.

4. Model Training: Trained Decision Tree Classifier with 80/20 train-test split.

5. Evaluation: Used confusion matrix and classification report to evaluate the model.

6. Visualization: Built Power BI dashboards showing age, gender, weekday trends, etc.

Conclusion

The project successfully demonstrated the application of machine learning in predicting healthcare appointment no-shows.
With the decision tree model, we identified key influencing factors such as waiting days, age, and SMS reminders.
Power BI dashboards enhanced interpretability by revealing hidden trends and enabling decision-makers to act on data-driven insights.
Going forward, integrating more advanced models like XGBoost and adding external factors like weather or transportation data could further improve prediction accuracy.
This solution offers a scalable approach for healthcare providers to optimize appointment scheduling and patient engagement strategies.