Healthcare Appointment No-Show Prediction

Abstract

Missed medical appointments are a major challenge in healthcare, leading to wasted resources, scheduling inefficiencies, and delays in patient treatment. This project aims to predict whether a patient will show up for their scheduled medical appointment using machine learning techniques. By analyzing various patient and appointment-related factors—such as age, gender, SMS reminders, waiting time, and neighborhood—the model helps healthcare providers anticipate no-shows and take preventive actions.

Introduction

In healthcare systems, patient no-shows significantly impact efficiency, staff utilization, and overall patient care quality. Hospitals and clinics face difficulties managing resources when patients fail to attend scheduled appointments. The objective of this project is to build a predictive model that can identify patients likely to miss their appointments. The insights from this model can help administrators send targeted reminders or adjust scheduling strategies to reduce missed visits.

Tools Used

- Programming Language: Python
- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- Environment: Jupyter Notebook / Google Colab
- Data Source: Kaggle – Medical Appointment No Shows Dataset

Steps Involved in Building the Project

1. Data Collection: The dataset containing patient appointment details was imported from Kaggle.
2. Data Cleaning: Removed missing and duplicate records, converted date columns, encoded categorical variables, and created new features such as waiting days.
3. Exploratory Data Analysis (EDA): Analyzed relationships, visualized distributions using histograms, boxplots, and heatmaps, and identified key factors.
4. Feature Engineering: Created derived variables, normalized data for model compatibility.
5. Model Building: Split data into training/testing sets, applied classification models like Logistic Regression, Random Forest, and Decision Tree, and evaluated performance.
6. Model Evaluation: Random Forest achieved the best results with high precision and recall; key predictors included waiting days, SMS reminders, and age.

Conclusion

The project successfully developed a machine learning model capable of predicting patient no-shows with good accuracy. The results demonstrate that appointment reminders and scheduling intervals play a critical role in attendance. By integrating such predictive tools, healthcare institutions can improve patient engagement, optimize resource allocation, and reduce wasted appointment slots. Future enhancements could include incorporating external

factors such as weather or traffic data for even more precise predictions.