Healthcare Appointment No-Show Prediction

Introduction

- No-shows are a major problem for hospitals, leading to wasted resources and delays in care. Predicting patient attendance can help optimize scheduling. Using the *appointments.csv* dataset, this project applies data analysis and machine learning to forecast no-shows and reveal factors affecting attendance.

Abstract

- Missed medical appointments cause inefficiency and financial loss in healthcare. This project predicts whether a patient will miss an appointment using machine learning. By analyzing patient and appointment data, key factors like waiting time, SMS reminders, and day of the week are identified to help reduce no-shows.

Tools Used

•**Python Libraries:** Pandas, NumPy, Matplotlib, Scikit-learn
•**Dataset:** `appointments.csv`
•**Techniques:** Data cleaning, visualization, feature engineering, and model training using classification algorithms.

Steps Involved in Building the Project

1.Data Cleaning: Removed duplicates, handled missing values, and formatted dates.
2.EDA: Explored trends in age, gender, waiting time, and SMS reminders.
3.Feature Engineering: Created new features and encoded categorical data.
4.Model Training: Used Logistic Regression, Decision Tree, and Random Forest.
5.Evaluation: Random Forest achieved the best accuracy and identified main predictors of no-shows.

Conclusion

The project successfully predicts patient no-shows using machine learning. Key factors like longer waiting times and lack of SMS reminders increase no-show risk. These insights can help healthcare providers send reminders or adjust scheduling, improving patient care and resource efficiency.