

Healthcare Appointment No-Show Prediction

1. Introduction

Missed healthcare appointments cause hospitals to lose money, increase patient wait times, and waste resources. By predicting which patients might miss their appointments, hospitals can improve scheduling and provide better care.

2. Abstract

This project is based on the "**Healthcare Appointment No-Show**" dataset, sourced from Kaggle. The dataset, containing patient appointment records, was analyzed to predict no-shows using machine learning techniques. Key patterns — including the impact of SMS reminders, waiting times, weekdays, and demographic factors — were carefully examined and visualized. A predictive model was developed and evaluated, and insights were presented through Power BI dashboards to support scheduling optimization strategies.

3. Tools Used

- **Python** (Pandas, Scikit-learn, XGBoost, Matplotlib, Seaborn)
- **Power BI** (Dashboard and Visualizations)

4. Steps Involved in Building the Project

Data Import and Cleaning

- Loaded appointment data (110,527 records) from Kaggle.
- Preprocessed data by renaming columns, handling datetime formats, creating new features (e.g., `waiting_days`, `appointment_weekday`), and encoding the target (`no_show`).

Exploratory Data Analysis (EDA)

- Analyzed trends such as:
 - SMS reminders reducing no-shows slightly.
 - Higher no-show rates among young adults.

- Certain weekdays (Monday, Friday) having more missed appointments.
- Longer waiting times correlating with more no-shows.

Model Building

- Trained multiple models:
 - **Decision Tree Classifier**
 - **Random Forest Classifier**
 - **XGBoost Classifier**
- Applied **SMOTE** oversampling to handle class imbalance.
- Evaluated models using confusion matrices and classification reports.
- Random Forest and XGBoost achieved higher recall (~80%) for no-show prediction, although overall accuracy was moderate (~59%).

Dashboard Development

- Built a Power BI dashboard showing:
 - **Total Appointments, No-Show Rate** (via cards)
 - **No-Show Rate by Weekday** (bar chart)
 - **No-Show Trend Over Time** (line chart)
 - **SMS vs No-Show Distribution** (pie chart)
 - Interactive filters for age, gender, scholarship status.

5. Conclusion

This project successfully demonstrated the ability to predict patient no-shows using machine learning. The Random Forest and XGBoost models were able to capture important patterns, although predicting no-shows remains a challenging task due to behavioral complexity. Power BI visualizations provided actionable insights for clinics to optimize scheduling — by reducing waiting days, improving reminder systems, and targeting high-risk patients for follow-ups.