Healthcare Appointment No-Show Prediction

My project Website link:- https://healthcare-appointment-no-show-prediction1.odoo.com/

1.Introduction

Healthcare systems around the world face many challenges, one of which is the issue of patients missing their scheduled appointments without prior notice. This phenomenon, known as 'no-show', affects resource utilization, reduces healthcare efficiency, and increases the waiting time for other patients. By leveraging data-driven approaches, it is possible to understand patterns in patient behavior and develop predictive models to forecast the likelihood of a no-show. This report details a machine learning project aimed at predicting no-shows based on historical appointment data.

---

1. Objective

The primary objective of this project is to develop a machine learning model that can predict whether a patient will show up for their medical appointment. This predictive system will assist healthcare providers in better planning, reducing no-show rates, and enhancing patient care delivery. Secondary objectives include performing thorough exploratory data analysis, identifying key features contributing to no-shows, and evaluating the performance of various classification models.

---

2. Dataset Description

The dataset used in this project is derived from over 100,000 medical appointments in Brazil and is publicly available on Kaggle. It includes patient demographics, appointment details, and whether the patient attended the appointment. The target variable is 'No-show', which is marked 'Yes' if the patient did not show up.

Key features in the dataset include:

- - PatientId, AppointmentID
- - Gender, Age
- - ScheduledDay, AppointmentDay
- - Neighbourhood
- - Scholarship (welfare program)
- - Hypertension, Diabetes, Alcoholism, Handcap
- - SMS_received
- - No-show (Target variable)
-

---

3. Tools and Technologies Used

The following tools and technologies were used in the development and analysis of this project:
- Python (programming language)
- Jupyter Notebook (IDE)
- Pandas and NumPy for data preprocessing
- Matplotlib and Seaborn for visualization
- Scikit-learn for model building

- XGBoost for boosting techniques
- Streamlit (optional) for deploying the model as a web application

---

4. Methodology

The approach followed in the project is outlined below:

1. 1. Data Cleaning and Preprocessing: Removing null values, converting date formats, and fixing incorrect data.
2. 2. Exploratory Data Analysis (EDA): Understanding the data distribution, trends, and correlations.
3. 3. Feature Engineering: Creating new relevant features such as waiting time, day of the week, etc.
4. 4. Model Training: Implementing classification algorithms including Logistic Regression, Random Forest, and XGBoost.
5. 5. Model Evaluation: Using accuracy, precision, recall, and F1-score to evaluate performance.
6. 6. Deployment: Optionally deploying the model using a simple user interface in Streamlit.

---

6. Model Evaluation

The performance of three models was compared. The table below summarizes the evaluation metrics:

Model	Accuracy	Precision	Recall	F1-score

Logistic Regression	79.3%	0.81	0.74	0.77
Random Forest	82.7%	0.83	0.78	0.80
XGBoost	84.1%	0.85	0.80	0.82

Among the models tested, XGBoost achieved the highest overall performance, making it the most suitable model for deployment. Its ability to handle imbalanced data and learn complex patterns contributed to its superior results.

---

7. Results and Discussion

The results indicate that certain features significantly impact the likelihood of a no-show. For instance, patients who received SMS reminders were more likely to attend. Other influential factors included waiting days between scheduling and appointment, patient age, and chronic conditions such as hypertension and diabetes. The imbalance in the dataset (more patients showed up than missed) was addressed during modeling to ensure fair evaluation.

---

8. Conclusion

This project demonstrated the effectiveness of machine learning in tackling real-world healthcare challenges. The predictive model built can assist healthcare providers in reducing missed appointments and improving the efficiency of clinical services. It also opens avenues for future research on behavioral patterns of patients.

---

9. Future Scope

- Integrate patient reminder systems with the predictive model.
- Extend the dataset with more patient history, time series analysis, and insurance data.
- Perform cost-benefit analysis for healthcare institutions based on predictions.
- Deploy the model in production using a web-based dashboard for hospital use.
- Apply advanced techniques such as deep learning for improved accuracy.