Introduction:

Healthcare systems often face challenges due to patients missing their scheduled appointments. Missed appointments can lead to inefficient use of medical resources, increased waiting times, and poor patient outcomes.

This project aims to predict whether a patient will show up for their medical appointment based on various features such as age, health conditions, and whether an SMS reminder was sent. The predictive insights can help healthcare providers improve appointment scheduling, resource planning, and patient engagement strategies.

Abstract:

The Healthcare Appointment No-Show Prediction project utilizes patient appointment data to develop a machine learning model that forecasts the likelihood of a patient missing their appointment.

Using Python (Pandas, Scikit-learn) for data preprocessing and model building, and Power BI for visualization, the project identifies patterns influencing attendance behaviour.

A Decision Tree Classifier was trained on features like age, medical conditions, and SMS reminders to predict the outcome. The model's performance was evaluated using metrics such as Accuracy, Precision, Recall, F1-score, MAE, and MSE.

The findings revealed that SMS reminders and patient demographics significantly impact attendance rates. The model can support data-driven strategies to minimize no-shows and optimize healthcare operations.

Tools Used:-

Tool / Library	Purpose
Python-	Core programming language for analysis and modelling
Pandas-	Data loading, cleaning, and manipulation
NumPy-	Numerical computation
Scikit-learn (sklearn)-	Machine learning model creation and evaluation
Matplotlib / Seaborn-	Data visualization and confusion matrix plotting
Power BI-	Visualization of model insights and trends
PyCharm-	Interactive environment for analysis and experimentation

Steps Involved in Building the Project:

1) Data Import and Exploration

- Loaded the dataset (Healthcare Appointment.csv) using Pandas.
- Checked for missing values, data types, and column distributions.

2) Data Cleaning and Preprocessing

- Renamed key columns for clarity (e.g., Showed_Up → target variable).
- Handled boolean and categorical data (e.g., SMS reminders).
- Converted the target column into binary format (Showed = 1 for attended, 0 for no-show).

3) Exploratory Data Analysis (EDA)

- Analyzed correlations between SMS reminders, age, and attendance rate.
- Visualized the no-show rate by SMS_received, age groups, and weekdays.

4) Feature Selection

- Selected relevant variables such as Age, Scholarship, Hipertension, Diabetes, Alcoholism, Handicap, and SMS.

5) Model Building

- Split data into training and testing sets (80:20 ratio).
- Trained a Decision Tree Classifier using sklearn.tree.

6) Model Evaluation

Evaluated using multiple metrics:

- Accuracy, Precision, Recall, and F1-score
- Mean Absolute Error (MAE) and Mean Squared Error (MSE)

Visualized the Confusion Matrix to analyze prediction distribution.

7) Visualization and Reporting

Created Power BI dashboards to display:

- Overall No- Shows
- Average age of patients
- Average waiting days between scheduled day and appointment day
- No-show trends by SMS reminders
- Showed up patients by gender

- Showed up patients by day of the week and day numbers
- Filters by SMS, Handicap, Diabetes, Alcoholism, Hipertension.

Compiled insights into actionable recommendations in the form of a pdf document.

Conclusion:

The project successfully demonstrates how machine learning can be applied in healthcare scheduling to reduce appointment no-shows.
The analysis indicates that sending SMS reminders significantly improves attendance rates.
Other factors like age and chronic conditions also contribute to patient reliability.
The Decision Tree model provides interpretable results that can guide administrative decisions.
By integrating these predictive insights into hospital management systems, healthcare providers can optimize scheduling, reduce idle time, and enhance patient care efficiency.