Post-Surgery Recovery Time Estimation Using Decision Tree

Project Title: Post-Surgery Recovery Time Estimation Using Decision Tree

1. Project Description

This project focuses on predicting the post-surgery recovery time for patients using a machine learning approach. The goal is to estimate how many days a patient will take to recover based on various input factors such as the type of surgery, age of the patient, and their pre-surgery vital signs (heart rate, blood pressure, and oxygen saturation). This helps hospitals and doctors in planning recovery times, improving care, and managing resources effectively.

2. Project Objective

- Build a Decision Tree Regression model to predict recovery time (in days).
- Use patient features like procedure type, age, and vitals to train the model.
- Provide accurate predictions to assist in medical planning and resource allocation.

3. Data Understanding

The dataset contains 1000 patient records with the following features:
- Age: Age of the patient.
- Procedure_Type: Type of surgery performed (e.g., Cardiac, Neuro, Plastic).
- Heart_Rate: Pre-surgery heart rate in bpm.
- Blood_Pressure: Pre-surgery blood pressure in mmHg.
- Oxygen_Saturation: Pre-surgery oxygen level in percentage.
- Recovery_Time_Days (Target): Number of days taken to recover.

4. Data Preprocessing

Before building the model, data preprocessing was performed to clean and transform the data:

- Separated features (X) and target (y) from the dataset.
- Identified categorical columns (Procedure_Type) for encoding.
- Used OneHotEncoder to convert categorical variables into numeric format.
- Applied ColumnTransformer to apply encoding only to relevant columns.
- Used train_test_split to divide data into training and testing sets (80% train, 20% test).

These steps ensured that the model received clean, numeric data suitable for training.

5. Model Selection and Training

A Decision Tree Regressor was selected due to its ability to handle both categorical and numerical features, model non-linear relationships, and its interpretability. The model was trained using the preprocessed training data. Hyperparameters like max_depth can be adjusted to optimize performance.

6. Model Evaluation

After training, predictions were made on the test set. Model performance was evaluated using:

- Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values.
- R^2 Score: Indicates how well the model explains the variance in the target variable.

These metrics provided insight into the accuracy and generalization capability of the model.

7. Data Visualization and Insights

- A histogram of recovery time showed most patients recover in 10–35 days.
- A boxplot revealed that cardiac and neuro surgeries generally take longer for recovery.
- A correlation heatmap showed positive correlation between age and recovery time, and slight negative correlation with vitals like oxygen saturation and heart rate.

These visualizations supported our assumptions and helped guide feature importance.

8. Final Conclusion

The Decision Tree model effectively predicted recovery time using patient features. It demonstrated that surgery type, age, and pre-surgery vitals significantly impact recovery duration. This model can help healthcare providers personalize recovery plans and better allocate medical resources.