Task 5: Train-Test Split & Evaluation Metrics

Tools:

- Python (Scikit-learn)

Dataset:

- "Heart Disease Dataset"

Hints / Mini Guide:

1. Split dataset into train and test sets.
2. Explain purpose of training vs testing.
3. Train a simple model (Logistic Regression).
4. Predict on test data.
5. Calculate accuracy, precision, recall.
6. Explain confusion matrix.
7. Interpret results.

Deliverables:

- Trained model
- Evaluation report

Final Outcome:
Intern understands model evaluation fundamentals.

Interview Questions Related To Above Task:

- What is overfitting?
- Accuracy vs precision?
- What is recall?
- What is confusion matrix?
- Why split data?

📌 **Task Submission Guidelines**

- ⏰ **Time Window:**

Youcan complete the task anytime between 10:00 AM to 10:00 PM on the given day. Submission link closes at 10:00 PM

- 🔍 **Self-Research Allowed:**

Youare free to explore, Google, or refer to tutorials to understand concepts and complete the task effectively.

- 🛠️ **Debug Yourself:**

Tryto resolve all errors by yourself. This helps you learn problem-solving and ensures you don't face the same issues in future tasks.

- 💸 **No Paid Tools:**

Ifthe task involves any paid software/tools, do not purchase anything. Just learn the process or find free alternatives.

- 📁 **GitHub Submission:**

Create a new GitHub repository for each task.

Add everything you used for the task — code, datasets, screenshots (if any), and a short README.md explaining what you did.

📤 **Submit Here:**

After completing the task, paste your GitHub repo link and submit it using the link below:

- 👉 [Submission Link]

⭐⭐⭐⭐⭐