✅ Task 15: End-to-End Machine Learning Pipeline

Dataset:
- Primary: Breast Cancer Dataset (sklearn)
- Alternative: Titanic Dataset

Tools:
- Python
- Pandas, NumPy
- Scikit-learn (Pipeline, ColumnTransformer)
- Jupyter Notebook / Google Colab

Hints / Mini Guide:

1. Load the dataset and separate features and target variable.
2. Identify numerical and categorical features for preprocessing.
3. Apply scaling and encoding using ColumnTransformer.
4. Create a complete ML pipeline combining preprocessing and model.
5. Split data into train and test sets.
6. Train the pipeline and generate predictions on test data.
7. Evaluate using accuracy, precision, recall, and F1-score.

Deliverables:

- Notebook with ML pipeline
- Evaluation metrics
- Saved pipeline model (.pkl)

Final Outcome:
Intern understands end-to-end ML workflow used in real production systems.

Interview Questions Related To Above Task:

- What is an ML pipeline?
- Why pipelines reduce data leakage?
- What is ColumnTransformer?
- Why pipelines help deployment?
- Pipeline vs manual preprocessing?

📌 **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]