**Task 4: Data Science Lifecycle Example**

**Application Selected:** **Fraud Detection in E-Commerce**

**1. Problem Definition**

- Online marketplaces face fraudulent activities like payment fraud, fake reviews, and account takeovers.

- The goal is to build a fraud detection system that identifies suspicious transactions in real time.

**2. Data Collection**

- Gather transaction data, user behaviour logs, and historical fraud cases.

- Data sources include customer purchases, login patterns, IP addresses, and device information.

**3. Data Cleaning & Preprocessing**

- Remove duplicates and manage missing values.

- Convert categorical data (e.g., payment method, location) into numerical format.

- Normalize transaction amounts to maintain consistency across different currencies.

**4. Model Training & Evaluation**

- Use machine learning models like Random Forest, Logistic Regression, or Neural Networks to classify transactions as fraudulent or legitimate.

- Split the dataset into training and testing sets.

- Evaluate performance using metrics like precision, recall, and F1-score.

**5. Deployment & Monitoring**

- Deploy the trained model into the e-commerce platform’s transaction processing system.

* Continuously monitor the model’s accuracy and update it with new fraud pattern.