Task 4: Data Science Lifecycle Example

Application: Healthcare (Disease Prediction)

1. **Problem Definition**
    - Define the goal: Predict the likelihood of patients developing a specific disease (e.g., diabetes) based on their medical history and lifestyle data.
2. **Data Collection**
    - Gather data from electronic health records (EHRs), wearable devices, and patient surveys. Include features like age, weight, blood pressure, and glucose levels.
3. **Data Cleaning & Preprocessing**
    - Handle missing values, remove duplicates, and normalize data. Encode categorical variables and split the dataset into training and testing sets.
4. **Model Training & Evaluation**
    - Train machine learning models (e.g., logistic regression, random forests) on the training data. Evaluate performance using metrics like accuracy, precision, and recall.
5. **Deployment & Monitoring**
    - Deploy the model into a healthcare system where doctors can input patient data to receive predictions. Continuously monitor the model's performance and retrain it with new data to maintain accuracy.