Week 2.1: Supervised Learning

Task 2.1: Building a Logistic Regression Model

Objective: Implement and evaluate a logistic regression model for binary classification.

Dataset: Heart Disease Dataset from UCI. This dataset includes patient data with various attributes like age, cholesterol levels, and heart rate, along with a target variable indicating the presence of heart disease.

- **Link to dataset**: [Heart Disease Dataset on UCI](https://archive.ics.uci.edu/ml/datasets/heart+disease) :
  https://archive.ics.uci.edu/ml/datasets/heart+disease

Activities:

1. **Data Preparation**:
   - Split the data into training and testing sets.
   - Ensure that the data is properly scaled or normalized if necessary.
2. **Model Building**:
   - Build a logistic regression model using scikit-learn.
   - Fit the model on the training data.
3. **Model Evaluation**:
   - Evaluate the model on the testing set using metrics such as accuracy, precision, recall, and the ROC curve.

Expected Output:

- A Jupyter notebook that includes the model building and evaluation process.
- Visualizations of the ***ROC curve and confusion matrix***.

Documentation:

- Follow the provided documentation template to outline the purpose, process, findings, and any insights from the exploratory data analysis.

General Guidelines for Tasks:

- **Comment your code**: Ensure your code in the Jupyter notebook is well-commented to explain why each step is performed.
- **Consistent Formatting**: Use clear headings and subheadings in your Jupyter notebooks and documentation.
- **Testing and Validation**: After each major step, use simple tests or checks to ensure the transformations are performed as expected.