## Week 1: Introduction to Machine Learning

---

### Task 1.2: Data Cleaning and Preparation

**Objective**: Learn the importance of preprocessing data to prepare for any machine learning model.

**Dataset**: **Titanic Dataset from Kaggle**. This dataset includes passenger information from the Titanic, such as age, fare, cabin, survival status, etc.

- **Link to dataset**: Titanic Dataset on Kaggle : https://www.kaggle.com/c/titanic

**Activities**:

1. **Load and Inspect the Dataset**:
    - Load the data into a pandas DataFrame.
    - Inspect the data for missing values, potential errors, and outliers.
2. **Data Cleaning**:
    - Handle missing values by filling them with the median or mode, or by using other appropriate imputation methods.
    - Remove outliers if necessary or treat them appropriately.
3. **Data Transformation**:
    - Convert categorical data into numeric format using one-hot encoding or label encoding.
    - Normalize or standardize the numerical values if required for later modeling.

**Expected Output**:

- A Jupyter notebook with detailed steps of data cleaning and preprocessing:
    - Before and after statistics of key columns.
    - Any transformations applied and the rationale behind these choices.

**Documentation**:

- Use the documentation template to detail every step taken in the preprocessing phase, including justifications for each choice (e.g., why certain outliers were removed or why specific columns were transformed).

**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.