



KHWAJA FAREED  
**UEIT**  
RAHIM YAR KHAN

Faculty of  
**Computer Sciences &  
Information Technology**

**Course Name:** Artificial Intelligence Lab

**Instructor:** Muhammad Amjad Raza

**Assignment Name:** Pre-processing Methods using sklearn.preprocessing

**Due Date:** 26-03-2023

**Total Marks:** 10

**Objective:** The objective of this assignment is to give you hands-on experience with some of the commonly used pre-processing methods in machine learning using the sklearn.preprocessing module.

## Tasks:

1. Load the dataset: Load the 'breast\_cancer' dataset from the sklearn.datasets module. The dataset contains information about breast cancer tumors and whether they are malignant or benign.
2. Split the data: Split the data into training and testing sets using the train\_test\_split function from the sklearn.model\_selection module.
3. You can download dataset from here <https://iotstuffs.com/breast-cancer-classification>
4. Preprocess the data:
  - Scale the data using the StandardScaler function from the sklearn.preprocessing module.
  - Encode the target variable using the LabelEncoder function from the sklearn.preprocessing module.
  - One-hot encode the target variable using the OneHotEncoder function from the sklearn.preprocessing module.
5. Build a model: Build a logistic regression model using the preprocessed data.
6. Evaluate the model: Calculate the accuracy, precision, recall, and F1 score of the logistic regression model on the test set.
7. Compare the results: Compare the results of the logistic regression model trained on the raw data with the results of the logistic regression model trained on the preprocessed data.

**Deliverables:** Submit a Jupyter notebook with the following:

1. Code to load and split the data
2. Code to preprocess the data
3. Code to build and evaluate the logistic regression model
4. A summary of the results, including a comparison of the results of the logistic regression model trained on the raw data with the results of the logistic regression model trained on the preprocessed data.



KHWAJA FAREED  
**UEIT**  
RAHIM YAR KHAN

Faculty of  
**Computer Sciences &  
Information Technology**

## Note:

1. You can use any other preprocessing methods available in sklearn.preprocessing module.
2. You can use any other classification model available in sklearn to compare the results of the logistic regression model.
3. You can use markdown cells in the Jupyter notebook to explain your approach and results.