# MCS2203 - Data Analytics and Machine Learning

## Assignment – 2

In this assignment, you are required to train a **supervised machine learning model** for the given data set using **any software tool** available.

In supervised learning, you have access to examples of correct input-output pairs that you can show to the machine during the training phase.

The dataset given has **699 samples** and the Training dataset consists of **490 samples** and the Test dataset consists of **209 samples**. There are only **two (2) classes** (called Class 2 & Class 4) that exist in the dataset. There are **nine (9) attributes** available in the dataset and all are numerical values. Class distribution of the Training and the Test datasets are as follows.

| **Training Dataset**: | **Test Dataset**: |
|---|---|
| Class 2: 315 samples | Class 2: 143 samples |
| Class 4: 175 samples | Class 4: 66 samples |

Further details of the dataset are available in the PGVLE in a file called "**MCS2203-Assignment-2-Details.txt**". You need to download the file called "**MCS2203-Assignment-2-Dataset.xlsx**" for the Training and the Test datasets.

a) Submit PowerPoint presentation slides including screenshots with an explanation of the tool you used for the above training process and its outputs.
b) Also, include the confusion matrix for the Test dataset after completing your training process for the Test dataset.

Confusion Metrix for Test dataset:

| | **Class 2 Actual** | **Class 4 Actual** |
|---|---|---|
| **Predicted as Class 2** | ?? | ?? |
| **Predicted as Class 4** | ?? | ?? |
| | **143** | **66** |

c) Calculate the following parameters for the Test dataset.
   I. Accuracy
   II. Precision
   III. Sensitivity
   IV. Specificity

Submit your answers in PowerPoint format and name your file starting with your index number.

**e.g. MCS2203-A2-xxxxxxxx.ppt (xxxxxxxx should be your index number)**

**Deadline**: Submit on or before **10th April 2020** to the PGVLE.

**Dr. Ajantha Atukorale**/ Course Lecturer