Instructions:

Please read the instructions very carefully.

1. This is an **INDIVIDUAL WORK** unless specified otherwise. Students are not allowed to share their answers (actual coding) but are allowed to discuss with each other to solve the problems.

2. The tasks will be checked during the practical session itself therefore students will be assessed based on participation and their answers.

3. To start:

    a. Create a new folder called **StudentID-P1**. Please change student id to your own personal ID and change the 1 to the correct practical number.

    b. Create one .ipynb / .py file for each question with the naming convention question1.ipynb, question2.ipynb and so on.

    c. The following information must be included in each file:
        i. Student Name
        ii. Student ID
        iii. Module Code and Title

4. Please note that **ALL FILE AND NAMING CONVENTIONS** must be followed.

5. The **GREEN** colored font in the sample output represents an input from the user.

6. The **BLUE** colored font in the sample output represents a dynamic output.

7. Please note that the colored fonts will vary on the values specified. In other words, they are just **SAMPLES** only.

8. All tasks must be completed within the session. Students are given enough time to complete the tasks listed.

9. Once completed, please create a zip file with the same name as your folder, and upload it to LMS before the end of the session. A submission link will be provided.

10. Students are encouraged to upload their work to their own GitHub account. Students are also encouraged to include the GitHub link in the submission.

11. Students are encouraged to ask questions during the practical if they encountered a problem.

Question 1:

Apply supervised modelling techniques on any datasets of your choosing. You are also required to apply any model evaluation for comparing the outcome of different supervised learning algorithm. The following algorithm must be implemented:

a) kNN
b) Naive Bayes
c) Decision Trees
d) Linear models – Logistic Regression
e) SVM

Sample Output:

Note:

1. *All accuracy is based on the test set and NOT training set.*
2. *The evaluation might be different, the sample might differ from your own result. It depends on what you set for your parameter. No restriction for the parameters. It is up to you to find the best ways to improve your model.*
3. *The dataset is divided into 80/20 spilt. 80% is training set and 20% as test set.*
4. *You are encouraged to also used cross-validation to build your model.*

The sample output will only show supervised modelling techniques applied on Iris dataset.

a) kNN:
Accuracy: 1.0
Recall: 1.0
Precision: 1.0
F1 Score: 1.0

Confusion Matrix:

11	0	0
0	13	0
0	0	6

b) Naïve Bayes:
Accuracy: 1.0
Recall: 1.0
Precision: 1.0
F1 Score: 1.0

Confusion Matrix:

11	0	0
0	13	0
0	0	6

c) Decision Trees
Accuracy: 0.966667

Confusion Matrix:

11	0	0
0	13	0
0	1	5

d) Logistic Regression
Accuracy: 0.93333

Confusion Matrix:

11	0	0
0	11	2
0	0	6

e) SVM:
Accuracy: 1.0
Recall: 1.0
Precision: 1.0
F1 Score: 1.0

Confusion Matrix:

11	0	0
0	3	0
0	0	6

- END OF FILE -