**PA 2: Classification - Assignment**

**Student Details**

Student Name : Puja Nitin Redij

ID: 1001651089

Notes: When submitting, fill your name and ID in this cell. Do not to forget to cite any external sources used by you. You will be submitting two different notebooks one for KNN, one for DT .

**Submission Instructions**

Rename this submission file as Example:

'yourLastName\_Last4digitsofyourID\_PA2\_KNN.ipynb',

'yourLastName\_Last4digitsofyourID\_PA2\_DT.ipynb',

Place this file inside the folder 'PA#2\_Classification'

**Programming Assignment Details**

Do not forget, this is an individual assignment!

For this assignment use Jupyter notebook, Panda, and scikit.

Load census dataset as provided.

Create a dataframe and print the first and last five records of your dataset. Print the class names.

Split your dataset 70% for training, and 30% for testing the classifier.

Use Euclidean distance for KNN.

Select only 2 attributes for training and testing your model.

Test it with three different number for neighbors and record the results. Add comments to explain your code and variable names.

For DT, use Gini and Entropy as your uncertainty metric

Derive all rules from the DT.

Show charts/graphs for your results.

Calculate

For all of them print the confusion matrix, and the classification Report (includes:precision, recall, f1-score, and support).

**Report**

For each classification task you need to submit a report (Microsoft Word, or PDF) that you have to:

describe the method that you use (e.g. Nearest Neighbors),

Describe the dataset, and explain how you selected the two attributes(for KNN)?

Visualize the classifier in a 2D projection, for all three different number of neighbors. Interpret and compare the results.

For DT, compare the results of Gini And Entropy

Do not to forget to cite your sources!