



# **IS4151/IS5451 – AIoT Solutions and Development**

## **AY 2024/25 Semester 2**

### **Practical Lab 08 – Machine Learning for IoT Data (II)**

#### **Part 1 – Basic Programming**

##### **PE08-1-1 – Clustering**

This exercise is based on the Forbes2000.csv dataset.

Clustering of sales, profits and assets – Build two k-means clustering models to cluster the world leading companies in the 2004 Forbes list using  $k = 2$  and  $k = 3$ . Which clustering model is the better one?

Calculate the within-cluster mean and standard deviation for each of the independent variables – sales, profits and assets – based on the better clustering model.

How do your findings compare with the supervised learning models that we have created earlier using regression (PE07-1-1 and PE07-1-2) and classification analysis (PE07-1-3)?

#### **Part 2 – Advanced Programming**

##### **PE08-2-1 – Predict Adult Census Income with Classification**

Recall that in PE07-2-2, you had predicted adult census income using a decision tree classifier.

Using the same dataset:

- a) Perform a logistic regression analysis and report your results and findings.
- b) Explain which classification techniques (decision tree or logistic regression) is better for analysing this dataset out of the two techniques that you have tried.