# Pass Task 1

Solve the following set of problems using Python and submit the code file with the extension .ipynb in Olympus as part of your pass activity.

Part 1

Load data from digitData2.csv file. The last column of each file presents the label and the rest of the columns are features.

1. Selecting the optimum k value using Silhouette Coefficient and plot the optimum k values.

2. Create clusters using Kmeans and Kmeans++ algorithms with optimal k value found in the previous problem. Report performances using appropriate evaluation metrics. Compare the results.

3. Now repeat clustering using Kmeans for 50 times and report the average performance. Again, compare the results that you have obtained in Q3 using Kmeans++ and explain the difference (if any).

4. Apply DBSCAN on this dataset (digitData2.csv) and find the optimum "eps" and "min_samples" value. Is the number of clusters the same as the cluster found in Q2? Explain the similarity or differences that you have found between two solutions.

Part 2

Load "diabetes" datasets from SKlearn and do the PCA for n=3, and plot the data.

1. What is the variance (%) covered by the first 3 components? How is this percentage of variance computed?

2. Is there any correlation between these three components?

3. Apply t-SNE on the same dataset, select the first 3 components and plot them.

4. Report the comparison between the results obtained using t-SNE and PCA

## Assessment feedback

The results with comments will be released within 5 business days from the due date.

## Referencing

You must correctly use the Harvard method in this assessment. See the Deakin referencing guide.

## Academic integrity, plagiarism and collusion

Plagiarism and collusion constitute extremely serious breaches of academic integrity. They are forms of cheating, and severe penalties are associated with them, including cancellation of marks for a specific assignment, for a specific unit or even exclusion from the course. If you are ever in doubt about how to properly use and cite a source of information refer to the referencing site above.

Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgement as to its authorship, the work of any other person or resubmits their own work from a previous assessment task.

Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose, with the intent of obtaining an advantage in submitting an assignment or other work.

Work submitted may be reproduced and/or communicated by the university for the purpose of assuring academic integrity of submissions: https://www.deakin.edu.au/students/study-support/referencing/academic-integrity.