

## Dhirubhai Ambani University (formerly DA-IICT)

# IT-582: Foundations of Machine Learning

## Assignment – 9: Unsupervised Learning I (k – means clustering)

**Question: - 1** Use [this link for dataset](#) (training set). Identify different classes present in the dataset. Perform the k-means clustering considering the given features:

- 'sepal length'.
- 'sepal length' and 'sepal width'.
- 'sepal length', 'sepal width' and 'petal length'.
- All features in the dataset.

Perform the k-means clustering by taking  $k = 4$ . Plot possible visualizations for each case. After getting the mean vectors in each scenario, classify (i.e., by checking given the datapoint and mean vectors for a particular scenario, to which mean vector it is nearest) the datapoints in the [test dataset](#). Plot the visualization of the classified datapoints.

**Question: - 2** Use sklearn to fetch the MNIST dataset. Flatten the image matrices. Use these flattened vectors to perform k-means clustering for  $k = 5$  to  $10$ . For every  $k$ , plot a bar-chart where x-axis depicts the cluster number and y-axis depicts the count of datapoints in that particular cluster. Now, for  $k = 5$  to  $k = 15$ , perform k-means clustering and plot the knee-elbow graph for these values.

**IF YOU REFER ANY MATERIAL FOR ANSWERING ANY QUESTIONS, PLEASE MENTION THEM IN THE LAST SECTION OF YOUR IPYNB UNDER SECTION NAMED 'REFERENCES'.**

**Open Assignment:** Make a website, where you upload a dataset and it gives a knee-elbow plot with different k-values.

[illegible][illegible]