# **HW Assignment 8: Unsupervised Learning (Dimension Reduction)** CS6140: Machine Learning Spring 2023

Due Date: Sunday, April 23, 2023  
**(10 points)**

# 

# Scenario

This assignment is a continuation of the second module on Unsupervised Learning. This assignment deals with the concept of dimension reduction. Dimension reduction is a technique to reduce the number of features or variables in a dataset while preserving its essential structure and relationships. It aims to simplify complex, high-dimensional data, making it easier to visualize, analyze, and process. Common dimension reduction techniques include Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation Projection (UMAP) and others. These methods (except PCA) work by projecting the high-dimensional data onto a lower-dimensional space, retaining as much relevant information as possible while removing noise and redundancy. Dimension reduction can improve computational efficiency and reduce overfitting.

# Tasks

You are given a sample program and a synthetically generated dataset.

The sample program demonstrates various dimension reduction techniques on a dataset. The sample program covers the following dimension reduction techniques: Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), Locally Linear Embedding (LLE), ISOMAP, and UMAP.

The synthetic data contains one hundred features, most of which are numeric. The numerical features are on different scales and some of them are associated / correlated with each other. There is also a categorical response variable.

Your task is to replicate code for the above five techniques on the given synthetic data along with visualizing the reduced dimensions and color coding them by the response column.

# **Expected Output**

**Please submit a fully executed jupyter notebook identifying question number and steps. Make sure to add comments to your solution.**