**Clustering & PCA Case Study with Wine Dataset**

**Faculty: Mahesh Anand. S**

1. Load the data (wine.xls) file and perform the following,
2. Refer to this link, <https://archive.ics.uci.edu/ml/datasets/wine> and read the description of the dataset and attach suitable headers.
3. Remove the first column (cultivator) and perform KMeans with (K=3) and show the 3D cluster plot with the most appropriate attributes in X,Y and Z –axis.
4. Remove the first column (cultivator) and perform Hierarchical Clustering (K=3) with the best linkage method and show the 3D cluster plot with the most appropriate attributes in X,Y and Z –axis.
5. Comment on the results obtained using KMeans and Hierarchical Clustering, which clustering method best suited for this dataset.
6. Apply PCA technique on the 13 independent variables and reduce the dimensions to (top-7) PC dimensions based on Eigen Values and reduce the original dataset to 178 x 7 dimension and fit a Supervised Learning model for (70%) of records and validate on (30%) of the records with precision and recall values.
7. Compare the above (e) results with fitting the same model with the original dataset (178 x 13) without applying PCA. [with 70% & 30% split]
8. Comment on the effect of PCA with reference to Bias – Variance trade off, between the model (e) and (f)
9. Perform KFold Cross Validation and check the consistency of the results across different classification model and comment on the validation results