

**ASSESSMENT 4**  
**BCSE209P MACHINE LEARNING LAB**  
**WINTER SEMESTER 2024 - 25**

- I. Implement the hierarchical clustering for Iris Flower dataset.
- II. Let us consider data set mushroom have 22 baseline variables like cap-shape, capcolor, odor and so on were obtained for each of  $n = 8124$  mushrooms, as well as the response of interest, a qualitative measure of category after baseline. Implement Principal Component Analysis for the given dataset. Perform the following tasks:

Tasks:

1. Data Loading and Exploration:
  - Load the provided dataset.
  - Explore the dataset to understand its characteristics, including the number of features and data points.
1. Data Preprocessing:
  - Standardize or normalize the data if necessary.
2. Covariance Matrix:
  - Calculate the covariance matrix of the dataset.
3. Eigenvalue and Eigenvector Computation:
  - Compute the eigenvalues and eigenvectors of the covariance matrix.
4. Explained Variance:
  - Calculate the explained variance for each principal component.
  - Determine the cumulative explained variance to decide on the number of principal components to retain.
5. Dimensionality Reduction:
  - Select the top 'k' principal components based on the cumulative explained variance.
  - Transform the data into the reduced-dimensional space.
6. Visualization:
  - Visualize the data in the reduced-dimensional space.
7. Interpretation:
  - Discuss the results and interpret the impact of dimensionality reduction on the dataset.

- III. Implement Random Forest Algorithm using the Bill authentication Dataset and evaluate its performance using appropriate metrics.
- IV. Implement Ada boost algorithm to classify flower species.
- V. Comparison of various Classification Machine Learning algorithms and evaluate its performance using the appropriate metrics.