**Walchand College of Engineering, Sangli**

## **Machine Learning Lab (6CS372)**

**TY BTech | AY 2023-2024 | Even Sem**

**Assignment 9**

**Clustering**

1. **Explore functions**
   1. Explore input and output parameters for following algorithms along with sample example: kMeans, Agglomerative, DBSCAN

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

* 1. Explore cluster quality metrics: silhouette, jaccard index, davis-bouldin.
  2. Explore elbow method.

1. **Apply clustering algorithms**
   1. Download dataset from <https://github.com/mwaskom/seaborn-data/blob/master/penguins.csv>
   2. Perform necessary visualizations and pre-processing steps. Drop the target column ‘Species’.
   3. Apply kMeans, AgglomerativeClustering and DBSCAN.
   4. Check how clusters can get visualized.
2. **Analysis:**
   1. Given that you have labelled data, how will you compare clusters formed by clustering scheme and classes in the data?
   2. Experiment with different values of k in kMeans and check its cluster quality metrics.
   3. Similarly, vary parameters of DBSCAN and AgglomerativeClustering algorithms and observe effects.

**Deliverables:**

Jupyter Notebook or Python script containing the implementation.

Report on summarizing the findings from experimentation and analysis in notebook itself. Any additional visualizations or insights gained during the experimentation. This can go in markdown cell.

**Note**: Ensure proper documentation and comments throughout the code to make it understandable.