

Assignment 3: Handling Imbalanced Datasets and Evaluating Classifier Performance

The purpose of this assignment is to explore techniques for addressing class imbalance in datasets, evaluate their effects, and compare classifier performance across original and balanced datasets.

Instructions:

1. Dataset Selection:

- Select an imbalanced dataset of your choice.
- Print and report the ratio of instances for each class in the dataset.

2. Resampling Techniques:

- Apply two undersampling techniques of your choice to balance the dataset.
- Apply two oversampling techniques of your choice to balance the dataset.

3. Train-Test Split and Class Ratios:

- For the original dataset and each of the balanced datasets, perform a train-test split (80-20).
- Print the ratio of classes for both the training set and the test set in each case.

4. Classification:

- Choose any classifier (e.g., Decision Tree, KNN, etc.) and train it on the training dataset from the original and balanced datasets.

- Evaluate the classifier's performance on the corresponding test dataset for each case.

5. Report Preparation:

- Provide a small description of the undersampling and oversampling techniques you used.
- Include the results of your experiments, highlighting:
 - Class ratios before and after resampling.
 - Class ratios for train and test splits in each case.
 - Performance metrics (e.g., accuracy, precision, recall, F1-score) of the classifier for each dataset.
- Summarize your findings, comparing the performance of the classifier on the original and balanced datasets.