

Assignment 4, Introduction to Machine Learning

In this assignment, you will handle class imbalance using techniques discussed in class. Refer to the Python notebook used to illustrate class imbalance. In the notebook, we illustrated different methods to deal with class imbalance with a dataset and KNN classification model.

In this assignment, you can refer to the same notebook and use the same data set. Below are the tasks you need to complete.

- Use any three of the five methods discussed in class to balance the training set.
- Instead of using KNN, use Decision Tree and Logistic Regression models. Note that you need to tune the parameters in these models.
- In the GridSearchCV, use the recall value for the minority class for the scoring criteria.
- After training on the balanced train set, display the classification report on the test set. Compare with the results from the model trained on the unbalanced training set.
- Also, compare the area under the precision-recall curve for the minority class in the testing set.