BNU-HKBU United International College
DS 4013: Data Mining (For DS students)
Fall 2022 Course Project 1: Classification Analysis

Description:
This is an individual project related to classification analysis. Given a dataset, the goal is to create an accurate classifier and make prediction on unseen records.

Submission Requirement:
Upon completion, each student must submit the following materials:

1. Test data and its prediction
2. Code
   a) You **MUST** implement the following models by yourself: KNN, Naïve Bayes and Perceptron.
   b) You **MUST** adopt **at least two** models besides the aforementioned three ones for your classification task. You do not need to implement by yourself, instead you can take advantage of open source libraries, for example *scikit-learn*.
   c) Your code must be executable without any bug and can read the test data to perform prediction and report the performance. Include a **README** file to introduce the information for your code and explain how to execute your code.
3. Implementation report
   In the report, the following components should be included:
   1. The workflow.
   2. The models adopted (your implementation as well as those provided by existing libraries)
   3. Experimental results of different models, e.g., Macro-averaging/Micro-averaging of precision/recall/F-score.
   4. Result analysis
      a) Which model achieves the **BEST** performance on this dataset? Why?
      b) Conduct error analysis for the models that do not perform well.

Assessment:

1. Classifier implementation and performance: 70%
2. Code: 10%
3. Project report: 20%
4. Bonus: 10%
   a) Novel strategy that improves the classifier performance, e.g., ensemble learning, data preprocessing methods, etc.
   b) Implementation of other models by yourself.