Assignment No. 5
EECS 690
Introduction to Machine Learning
Due: 11:59 PM, Friday, October 30, 2020
Submit deliverables in a single zip file to BlackBoard
Name of the zip file: FirstnameLastname_Assignment5
Name of the Assignment folder within the zip file: FirstnameLastname_Assignment5

Deliverables:

1. Copy of Rubric5.docx with your name and ID filled out (do not submit a PDF)
2. Python source code.
3. Screen print showing the successful execution of your Python code. (Copy and paste the output from the Python console screen to a Word document and PDF it).

Assignment:

- I have created an imbalanced iris dataset called "imbalanced iris.csv" to use for this assignment which is located in the Assignment 5 folder.
- Use 2-fold cross-validation using the Neural Network machine learning model for each part.
- This assignment has three parts (listed below by number). Before each part's printout, printout the part number.

Part 1: Imbalanced Data Set

- Print out and label the Confusion Matrix and Accuracy score using the imbalanced iris data set.
- Print out and label the calculated Class Balanced Accuracy as described on Slide 16 of the 10/12 Lecture.
- Print out and label the calculated Balanced Accuracy as described on Slide 16 of the 10/12 Lecture.
- Print out and label the balanced accuracy score calculated by the skikit-learn function balanced_accuracy_score as described on Slide 20 of the 10/12 lecture.

Part 2: Oversampling

- Use the scikit-learn imblearn.over_sampling library (Slide 11 of the 10/14 lecture) for each of the following:
    - Balance the imbalanced iris dataset with random oversampling and print-out and label the Confusion Matrix and Accuracy score. (We don't need to use one of the balanced accuracy scores from Part 1 because the set is balanced now).
    - Balance the imbalanced iris dataset with SMOTE oversampling and print-out and label the Confusion Matrix and Accuracy score.
    - Balance the imbalanced iris dataset with ADASYN oversampling and print-out and label the Confusion Matrix and Accuracy score.

Part 3: Undersampling

- Use the scikit-learn imblearn.under_sampling library (Slide 11 of 10/16 lecture) for each of the following:
    - Balance the imbalanced iris dataset with random undersampling and print-out and label the Confusion Matrix and Accuracy score.

    - Balance the imbalanced iris dataset with Cluster undersampling and print-out and label the Confusion Matrix and Accuracy score.
    - Balance the imbalanced iris dataset with Tomek links undersampling and print-out and label the Confusion Matrix and Accuracy score.

Remember:

- Your Programming Assignments are individual-effort.
- You can brainstorm with other students and help them work through problems in their programs, but everyone should have their own unique assignment programs.