**CST8390 - Lab 2**

**K Nearest Neighbor (kNN)**

Name: Wei-Ting Yang

Student Number:040852642

1. 1. What is the **percentage** of correctly classified items? \_\_\_94.9438%\_\_\_\_\_\_
   2. What is the Weighted average of True Positive (TP) rate?\_\_\_\_\_\_0.949\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. Look at the confusion matrix, which class is incorrectly classified?

\_\_\_Class b\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* 1. What is the percentage of correctly classified instances? \_\_\_\_\_\_\_94.9438%\_\_\_\_\_
  2. What is the Weighted average of True Positive (TP) rate?\_\_\_\_\_0.949\_\_\_\_\_\_\_\_\_\_
  3. Look at the confusion matrix, which classes are incorrectly classified? Class b

1. Fill in the table.

|  |  |  |
| --- | --- | --- |
| K | Percentage of correctly classified instances | Number of instances misclassified in each class |
| 3 | 94.9438% | 1:0  2:9  3:0 |
| 5 | 95.5056% | 1:0  2:7  3:1 |
| 9 | 96.0674% | 1:0  2:6  3:1 |
| 11 | 97.191% | 1:0  2:5  3:0 |

1. Repeat step 9 with “Percentage Split” of 70. Fill in the following table.

|  |  |  |
| --- | --- | --- |
| K | Percentage of correctly classified instances | Number of instances misclassified in **each** class |
| 3 | 100% | 1:0  2:0  3:0 |
| 5 | 98.1132% | 1:0  2:1  3:0 |
| 9 | 100% | 1:0  2:0  3:0 |
| 11 | 100% | 1:0  2:0  3:0 |

1. Explanation of the process and the screenshot.

Student Number: 020452642

Total number of instances: 178

Student Number % Total number of instances: 146

Instance to be tested:146

Steps of the Test Process:

=SQRT((12.25-B146)^2+(3.88-C146)^2+(2.2-D146)^2+(18.5-E146)^2+(112-F146)^2+(1.38-G146)^2+(0.78-H146)^2+(0.29-I146)^2+(1.14-J146)^2+(8.21-K146)^2+(0.65-L146)^2+(2-M146)^2+(855-N146)^2))

1.Find the test instance number by taking the modulus of the student number with the total number of instances. Then, locate and select the corresponding row in the dataset.

2.Determine the distance between the target instance and all other instances using the provided calculation.

3.Create a new sheet and transfer the class attribute (column) and the calculated distance column onto it. Arrange the data by sorting it in ascending order based on distance.

4.Identify the top 9 entries with the smallest distances and tally the occurrence of each class number. The predominant class number among these entries signifies the class of the target instance.

A screenshot of a computer

Description automatically generated

1. Paste a screenshot of the histogram.

A screenshot of a computer

Description automatically generated