**CST8390 - Lab 3**

**K Nearest Neighbor (kNN)**

Name: Onur Onel

Student Number: 041074824

* 1. What is the **percentage** of correctly classified items? **94.9438 %**
  2. What are the True Positive (TP) rates of **each** class?

**1.000, 0.873, 1.000**

Look at the confusion matrix, which class is incorrectly classified?

**5 62 4 | b = 2**

|  |  |  |
| --- | --- | --- |
| K | Percentage of correctly classified instances | Number of instances misclassified in each class |
| 3 | 94.9438 % | 1: 59 0 0 | a = 1  2: 4 62 5 | b = 2  3: 0 0 48 | c = 3 |
| 5 | 95.5056 % | 1: 59 0 0 | a = 1  2: 2 64 5 | b = 2  3: 0 1 47 | c = 3 |
| 7 | 94.9438 % | 1: 59 0 0 | a = 1  2: 3 63 5 | b = 2  3: 0 1 47 | c = 3 |
| 9 | 96.0674 % | 1: 59 0 0 | a = 1  2: 2 65 4 | b = 2  3: 0 1 47 | c = 3 |

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: 19 0 0 | a = 1  2: 0 16 0 | b = 2  3: 0 0 18 | c = 3 |
| 5 | 98.1132 % | 1: 19 0 0 | a = 1  2: 0 15 1 | b = 2  3: 0 0 18 | c = 3 |
| 7 | 100 % | 1: 19 0 0 | a = 1  2: 0 16 0 | b = 2  3: 0 0 18 | c = 3 |
| 9 | 100 % | 1: 19 0 0 | a = 1  2: 0 16 0 | b = 2  3: 0 0 18 | c = 3 |

1. Explanation of the process and the screenshot.

**Calculated the Euclidean distance between the test instance and each training instance. In the test data sheet, the Euclidean distance between the test instance (attributes B2 to N2) and each training instance (attributes B3 to N\_last\_row) using the following formula**

**=SQRT(SUM((B2-N2)^2:(N2-N\_last\_row)^2))**



