**CST8390 - Lab 3**

**K Nearest Neighbor (kNN)**

Name: Neil Kingdom

Student Number: 040967309

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

Class 1 = 1.000

Class 2 = 0.873

Class 3 = 1.000

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

Class 2/b is incorrect. It classified a count of 62, not 71.

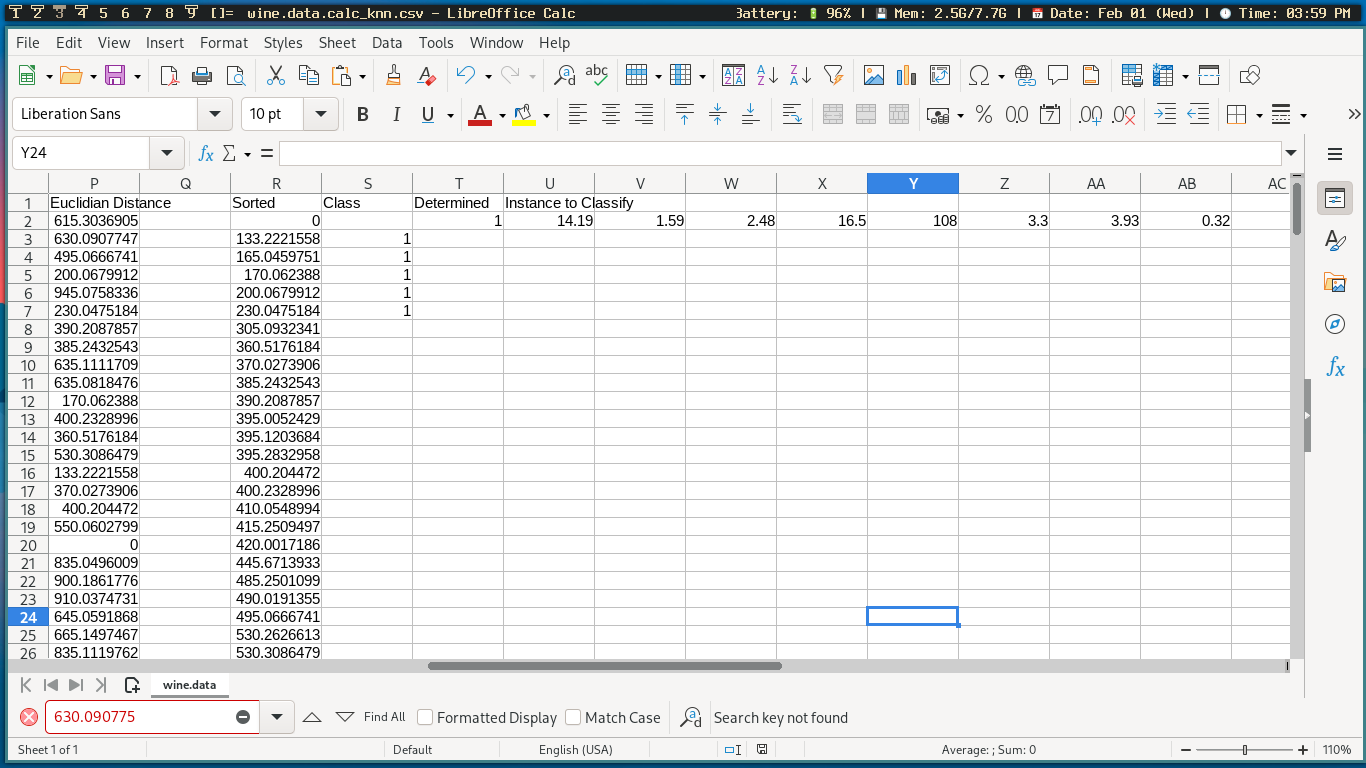
1. Fill in the table.
2. Fill in the table.

|  |  |  |
| --- | --- | --- |
| 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 |
| 7 | 94.9438% | 1: 0  2: 8  3: 1 |
| 9 | 96.0674% | 1: 0  2: 6  3: 1 |

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 |
| 7 | 100% | 1: 0  2: 0  3: 0 |
| 9 | 100% | 1: 0  2: 0  3: 0 |

1. Explanation of the process and the screenshot.

  
Offscreen is the dataset within the same Excel file. The Euclidian distance is calculated by taking the pythegorean theorum between the instance that we are trying to find the classification for and the current row of the dataset. We do this for each row of the dataset, calculating each distance between the current data point and the instance that we are trying to find. Then, we sort the distances from least to greatest. The class is determined by finding the row that the sorted number belonged to in the Euclidian Distance column and then getting its associated class. By taking an average of the first ‘k’ rows (in our case, 5), and then rounding up to remove any remainders, we get the class that the instance we are looking for most likely belongs to!