**kNN TigerFish Report**

**Dane Acena**

# **Problem Description**

Given a data set from the Clemson Wildlife and Fisheries Biology graduate students of two newly discovered species of fish in Lake Hartwell, develop a kNN classification algorithm that will determine a fish is a TigerFish1 (the positive case) or a TigerFish0.

# **Data Description**

The initial data consisted of 300 records representing features of either TigerFish1 species or TigerFish0 species with three tab-delimited entries. The first two are floats indicating the measured body length and dorsal fin length of each fish respectively. The last element is a digit, either “1” or “0” identifying the species of fish as either “TigerFish1” or “TigerFish0”. A plot of the initial dataset is shown in Figure 1.

![A screenshot of a cell phone

Description automatically generated]()

Figure 1: The Initial Data Set

# **Training a kNN Algorithm**

Dfd

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 |
| Test1 Errors |  |  |  |  |  |  |  |  |  |  |  |
| Test2 Errors |  |  |  |  |  |  |  |  |  |  |  |
| Test3 Errors |  |  |  |  |  |  |  |  |  |  |  |
| Test4 Errors |  |  |  |  |  |  |  |  |  |  |  |
| Test5 Errors |  |  |  |  |  |  |  |  |  |  |  |
| TOTALS |  |  |  |  |  |  |  |  |  |  |  |

Figure 2: Misclassifications for different values of k on the five training sets

# **A screenshot of a cell phone Description automatically generated**

Figure 3: Average accuracy for different values of k (PLACEHOLDER)

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Predicted TigerFish1 | |
| N | Y |
| Actual  TigerFish0 | N | TN = | FP = |
| Y | FN = | TP = |

Figure 4: Confusion Matrix

# **Results**

![A screenshot of a cell phone

Description automatically generated]()