Abir Razzak (amr439@drexel.edu)

INFO371 – Data Mining Applications

Assignment 1

Please review the .arff file for more information on the dataset.

# Weka Log

13:32:42: Weka Explorer

13:32:42: (c) 1999-2018 The University of Waikato, Hamilton, New Zealand

13:32:42: web: http://www.cs.waikato.ac.nz/~ml/weka/

13:32:42: Started on Sunday, 6 October 2019

13:33:08: Base relation is now PulsarStar (20 instances)

13:34:45: Started weka.classifiers.lazy.IBk

13:34:45: Command: weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

13:34:45: Finished weka.classifiers.lazy.IBk

# Weka Output

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: PulsarStar

Instances: 20

Attributes: 3

STD Integrated Profile

STD DM-SNR Curve

Pulsar

Test mode: split 80.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier

using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Predictions on test split ===

inst# actual predicted error prediction

1 1:0 1:0 0.944

2 2:1 2:1 0.944

3 1:0 1:0 0.944

4 2:1 2:1 0.944

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances 4 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0.0556

Root mean squared error 0.0556

Relative absolute error 11.1111 %

Root relative squared error 10.8465 %

Total Number of Instances 4

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 0

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 1

Weighted Avg. 1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000

=== Confusion Matrix ===

a b <-- classified as

2 0 | a = 0

0 2 | b = 1

# Data Plot

A close up of a white wall

Description automatically generated

*Legend*

Green – Pulsar Star

Red – Not Pulsar Star

Black – Testing Data (last 4 instances)

*Plotted in Excel*

*Manually Examining 3 Nearest Neighbors*

A close up of a map

Description automatically generated

*(Test data points are numbered in order from highest y value to lowest y value, that is to say from top to bottom)*

Point 1

* Closest Neighbor: Green
* 3 Closest Neighbors: 3 Green
* No discrepancy between k=1 and k=3
* ACTUAL DATA: Green (Pulsar)

Point 2

* Closest Neighbor: Green
* 3 Closest Neighbors: 3 Green
* No discrepancy between k=1 and k=3
* ACTUAL DATA: Green (Pulsar)

Point 3

* Closest Neighbor: Red
* 3 Closest Neighbors: 3 Red
* No discrepancy between k=1 and k=3
* ACTUAL DATA: Red (Not Pulsar)

Point 4

* Closest Neighbor: Red
* 3 Closest Neighbors: 2 Red, 1 Green
* No discrepancy between k=1 and k=3
* ACTUAL DATA: Red (Not Pulsar)