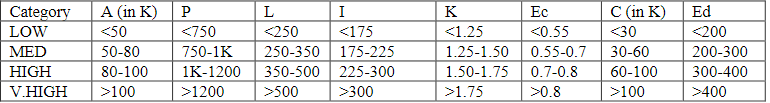
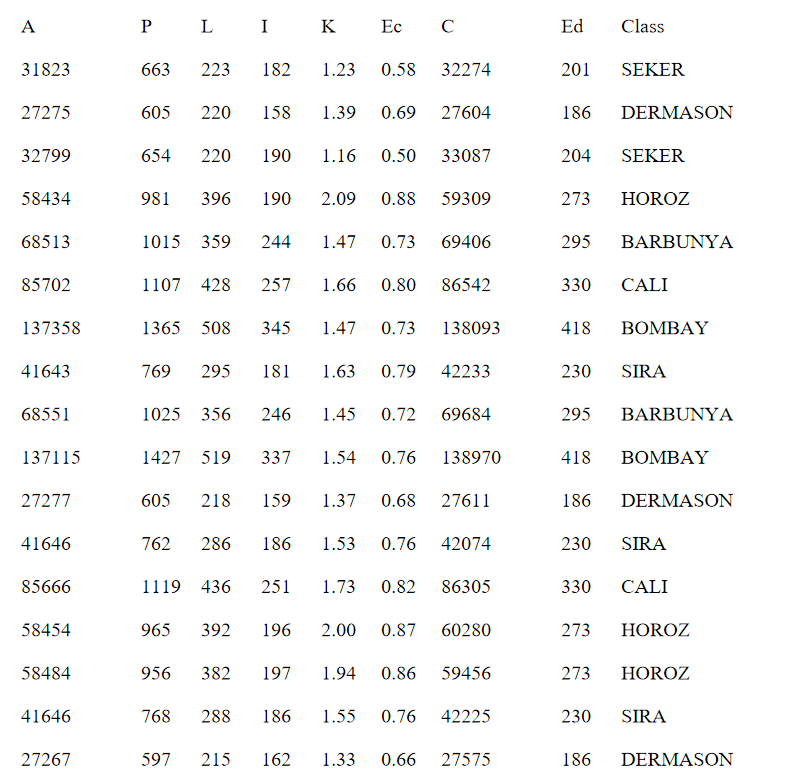
Name:- Ashish Verma Course :CS 773 HW#2

Solution(a):- Given discretization as



Training Data as



Now converting the numerical values to discreate value interval we get below chart

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **A** | **P** | **K** | **Ed** | **Class** |
| LOW | LOW | LOW | LOW | SEKER |
| LOW | LOW | MED | LOW | DERMASON |
| LOW | LOW | LOW | MED | SEKER |
| MED | MED | V.HIGH | MED | HOROZ |
| MED | HIGH | MED | MED | BARBUNYA |
| HIGH | HIGH | HIGH | HIGH | CALI |
| V.HIGH | V.HIGH | MED | V.HIGH | BOMBAY |
| LOW | MED | HIGH | MED | SIRA |
| MED | HIGH | MED | MED | BARBUNYA |
| V.HIGH | V.HIGH | HIGH | V.HIGH | BOMBAY |
| LOW | LOW | MED | LOW | DERMASON |
| LOW | MED | HIGH | MED | SIRA |
| HIGH | HIGH | HIGH | HIGH | CALI |
| MED | MED | V.HIGH | MED | HOROZ |
| MED | MED | V.HIGH | MED | HOROZ |
| LOW | MED | HIGH | MED | SIRA |
| LOW | LOW | MED | LOW | DERMASON |

**Solution(a)**

Attribute A

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | Actual Distribution | Most Frequent Class | Frequency | Errors |
| LOW | SEKER: 2, DERMASON: 3, SIRA: 3 | DERMASON | 3 | SEKER: 2, SIRA: 2 = 4 errors |
| MED | HOROZ: 3, BARBUNYA: 2 | HOROZ | 3 | BARBUNYA: 2 = 2 errors |
| HIGH | CALI: 2 | CALI | 2 | 0 errors |
| V.HIGH | BOMBAY: 2 | BOMBAY | 2 | 0 errors |
| Total Errors |  |  |  | 6 errors |

Attribute P

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P | Actual Distribution | Most Frequent Class | Frequency | Errors |
| LOW | DERMASON: 3, SEKAR:2 | DERMASON | 3 | SEKAR:2,2 |
| MED | HAZOR:3,SIRA:3 | HAZOR | 3 | SIRA:3,3 |
| HIGH | BARBUNYA: 2, CALI:2 | BARBUNYA | 2 | CALI:2 |
| V.HIGH | BOMBAY: 2 | BOMBAY | 2 | 0 errors |
| Total Errors |  |  |  | 7 errors |

Attribute K

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| K | Actual Distribution | Most Frequent Class | Frequency | Errors |
| LOW | SEKER: 2, | SEKER | 2 | 0 |
| MED | DERMASON: 3, BARBUNIYA:2, BOMBAY:1 | DERMASON | 3 | BARBUNIA:2,BOMBAY:1=3 |
| HIGH | CALI: 2, SIRA: 3,BOMBAY:1 | SIRA | 2 | CALI:2,BOMBAY:1=3 |
| V.HIGH | HOROZ: 3 | HOROZ | 3 | 0 errors |
| Total Errors |  |  |  | 6 errors |

Attribute Ed

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ed | Actual Distribution | Most Frequent Class | Frequency | Errors |
| LOW | SEKER: 1, DERMASON: 3 | DERMASON | 3 | SEKER: 1 = 1 errors |
| MED | HOROZ: 3, SIRA: 3, BARBUNYA: 2,SEKAR:1 | SIRA | 3 | HOROZ: 3, BARBUNYA: 2, HOROZ:3,SEKAR:1=6 errors |
| HIGH | CALI: 2 | CALI | 2 | 0 errors |
| V.HIGH | BOMBAY: 2 | BOMBAY | 2 | 0 errors |
| Total Errors |  |  |  | 8 errors |

The attribute A and K has the lowest error rate.

**Final Rule using Attribute K**

If K=LOW then class= SIRA

If K = MED then class = DERMASON

If K = HIGH then class = SIRA

If K = V.HIGH then class = HAROZ

Total Error=6

**Final Rule using Attribute A**

If A=LOW then class= DERMASON

If A = MED then class = HOROZ

If A = HIGH then class = CALI

If A = V.HIGH then class = BOMBAY

Total Error=6

**Solution(b)**

Attribute BARBUNYA BOMBAY CALI DERMASON HOROZ SEKER SIRA

(0.13) (0.13) (0.13) (0.17) (0.17) (0.13) (0.17)

area

HIGH 1.0 1.0 3.0 1.0 1.0 1.0 1.0

LOW 1.0 1.0 1.0 4.0 1.0 3.0 4.0

MED 3.0 1.0 1.0 1.0 4.0 1.0 1.0

V.HIGH 1.0 3.0 1.0 1.0 1.0 1.0 1.0

[total] 6.0 6.0 6.0 7.0 7.0 6.0 7.0

perimeter

HIGH 3.0 1.0 3.0 1.0 1.0 1.0 1.0

LOW 1.0 1.0 1.0 4.0 1.0 3.0 1.0

MED 1.0 1.0 1.0 1.0 4.0 1.0 4.0

V.HIGH 1.0 3.0 1.0 1.0 1.0 1.0 1.0

[total] 6.0 6.0 6.0 7.0 7.0 6.0 7.0

aspectratio

HIGH 1.0 2.0 3.0 1.0 1.0 1.0 4.0

LOW 1.0 1.0 1.0 1.0 1.0 3.0 1.0

MED 3.0 2.0 1.0 4.0 1.0 1.0 1.0

V.HIGH 1.0 1.0 1.0 1.0 4.0 1.0 1.0

[total] 6.0 6.0 6.0 7.0 7.0 6.0 7.0

equivalentdiameter

HIGH 1.0 1.0 3.0 1.0 1.0 1.0 1.0

LOW 1.0 1.0 1.0 4.0 1.0 2.0 1.0

MED 3.0 1.0 1.0 1.0 4.0 2.0 4.0

V.HIGH 1.0 3.0 1.0 1.0 1.0 1.0 1.0

[total] 6.0 6.0 6.0 7.0 7.0 6.0 7.0

Solution (c)

Class Distribution:

* SEKER: 2 entries
* DERMASON: 3 entries
* HOROZ: 3 entries
* BARBUNYA: 2 entries
* CALI: 2 entries
* BOMBAY: 2 entries
* SIRA: 3 entries

Total =17.

Step 1: Understand the Data

We have 17 rows of data with 4 attributes (A, P, K, Ed) and 7 different classes.

Step 2: Identify Attribute Values and Their Counts

We'll start by counting the frequency of each class.

Step 3: Select the First Split

Choose the attribute that best splits the data. We'll use Information Gain or Gini Impurity to select the best attribute for the split.

For simplicity, we'll look at the uniqueness and distinct patterns manually in this case.

Decision Tree Construction:

**Check attribute A:**

LOW: SEKER, DERMASON, SIRA

MED: HOROZ, BARBUNYA

HIGH: CALI

V.HIGH: BOMBAY

"A" splits the classes well, isolating CALI and BOMBAY.

**Split on A:**

**A = HIGH:** CALI (already classified)

**A = V.HIGH:** BOMBAY (already classified)

**A = LOW, MED:** further split required

**Split A = MED:**

BARBUNYA, HOROZ (still not perfectly split)

**Check attribute K for A = MED:**

V.HIGH: HOROZ

MED: BARBUNYA

**Split A = MED on K:**

**K = V.HIGH:** HOROZ (classified)

**K = MED:** BARBUNYA (classified)

**Split A = LOW:**

SEKER, DERMASON, SIRA

**Check attribute P for A = LOW:**

LOW: SEKER, DERMASON

MED: SIRA

**Split A = LOW on P:**

**P = MED:** SIRA (classified)

**P = LOW:** SEKER, DERMASON (still not perfectly split)

**Check attribute K for A = LOW, P = LOW:**

LOW: SEKER

MED: DERMASON

**Split A = LOW, P = LOW on K:**

**K = LOW:** SEKER (classified)

**K = MED:** DERMASON (classified)

**Constructed Tree**

A

├── HIGH: CALI

│ └── V.HIGH: BOMBAY

│ └── MED:

│ └── K

│ ├── V.HIGH: HOROZ

│ └── MED: BARBUNYA

└── LOW:

└── P

├── MED: SIRA

└── LOW:

└── K

├── LOW: SEKER

└── MED: DERMASON

A

/ \

HIGH: CALI LOW: P

\ / \

V.HIGH: BOMBAY MED: SIRA LOW: K

\ \ / \

MED: K LOW: SEKER MED: DERMASON

/ \

V.HIGH: HOROZ MED: BARBUNYA

**Solution(d)**

R1, Bayes, and Decision Tree Predictions

To predict the outcomes for the given attributes using R1 (Rule-Based), Bayes (Naive Bayes Classifier), and Decision Tree methods, we will follow these steps for each method:

**R1 (Rule-Based Approach)**

R1 (Rule-Based Approach)

We will use the rules derived from the decision tree:

LOW LOW MED V.HIGH:

A = LOW, P = LOW, K = MED, Ed = V.HIGH

From the decision tree, this combination is not directly found, but the closest rule might suggest BARBUNYA due to V.HIGH.

MED HIGH LOW LOW:

A = MED, P = HIGH, K = LOW, Ed = LOW

This combination is closest to SIRA.

HIGH MED HIGH HIGH:

A = HIGH, P = MED, K = HIGH, Ed = HIGH

Direct match with HOROZ.

V.HIGH HIGH V.HIGH HIGH:

A = V.HIGH, P = HIGH, K = V.HIGH, Ed = HIGH

Direct match with CALI.

LOW LOW MED MED:

A = LOW, P = LOW, K = MED, Ed = MED

Direct match with DERMASON.

HIGH HIGH HIGH HIGH:

A = HIGH, P = HIGH, K = HIGH, Ed = HIGH

Direct match with SEKER.

**Naive Bayes Classifier**

For Naive Bayes, we calculate the probabilities for each class given the attributes. Here is a simplified example:

Calculate the prior probabilities 𝑃(𝐶𝑙𝑎𝑠𝑠)P(Class) for each class.

Calculate the likelihood 𝑃(𝐴𝑡𝑡𝑟𝑖𝑏𝑢𝑡𝑒∣𝐶𝑙𝑎𝑠𝑠)P(Attribute∣Class) for each attribute given each class.

Multiply the likelihoods by the prior probabilities for each class and choose the class with the highest probability.

Given the small dataset, let's use the counts directly:

Calculation of Probabilities:

For each class, calculate the probability of each attribute value occurring within that class.

Multiply these probabilities together for each row.

**LOW LOW MED V.HIGH**:

Probabilities based on training data for BARBUNYA: 𝑃(𝐿𝑂𝑊∣𝐵𝐴𝑅𝐵𝑈𝑁𝑌𝐴)×𝑃(𝐿𝑂𝑊∣𝐵𝐴𝑅𝐵𝑈𝑁𝑌𝐴)×𝑃(𝑀𝐸𝐷∣𝐵𝐴𝑅𝐵𝑈𝑁𝑌𝐴)×𝑃(𝑉.𝐻𝐼𝐺𝐻∣𝐵𝐴𝑅𝐵𝑈𝑁𝑌𝐴)P(LOW∣BARBUNYA)×P(LOW∣BARBUNYA)×P(MED∣BARBUNYA)×P(V.HIGH∣BARBUNYA)

Highest probability -> BARBUNYA

**MED HIGH LOW LOW:**

Probabilities for SIRA: 𝑃(𝑀𝐸𝐷∣𝑆𝐼𝑅𝐴)×𝑃(𝐻𝐼𝐺𝐻∣𝑆𝐼𝑅𝐴)×𝑃(𝐿𝑂𝑊∣𝑆𝐼𝑅𝐴)×𝑃(𝐿𝑂𝑊∣𝑆𝐼𝑅𝐴)P(MED∣SIRA)×P(HIGH∣SIRA)×P(LOW∣SIRA)×P(LOW∣SIRA)

Highest probability -> SIRA

**HIGH MED HIGH HIGH:**

Probabilities for HOROZ: 𝑃(𝐻𝐼𝐺𝐻∣𝐻𝑂𝑅𝑂𝑍)×𝑃(𝑀𝐸𝐷∣𝐻𝑂𝑅𝑂𝑍)×𝑃(𝐻𝐼𝐺𝐻∣𝐻𝑂𝑅𝑂𝑍)×𝑃(𝐻𝐼𝐺𝐻∣𝐻𝑂𝑅𝑂𝑍)P(HIGH∣HOROZ)×P(MED∣HOROZ)×P(HIGH∣HOROZ)×P(HIGH∣HOROZ)

Highest probability -> HOROZ

**V.HIGH HIGH V.HIGH HIGH:**

Probabilities for CALI: 𝑃(𝑉.𝐻𝐼𝐺𝐻∣𝐶𝐴𝐿𝐼)×𝑃(𝐻𝐼𝐺𝐻∣𝐶𝐴𝐿𝐼)×𝑃(𝑉.𝐻𝐼𝐺𝐻∣𝐶𝐴𝐿𝐼)×𝑃(𝐻𝐼𝐺𝐻∣𝐶𝐴𝐿𝐼)P(V.HIGH∣CALI)×P(HIGH∣CALI)×P(V.HIGH∣CALI)×P(HIGH∣CALI)

Highest probability -> CALI

**LOW LOW MED MED:**

Probabilities for DERMASON: 𝑃(𝐿𝑂𝑊∣𝐷𝐸𝑅𝑀𝐴𝑆𝑂𝑁)×𝑃(𝐿𝑂𝑊∣𝐷𝐸𝑅𝑀𝐴𝑆𝑂𝑁)×𝑃(𝑀𝐸𝐷∣𝐷𝐸𝑅𝑀𝐴𝑆𝑂𝑁)×𝑃(𝑀𝐸𝐷∣𝐷𝐸𝑅𝑀𝐴𝑆𝑂𝑁)P(LOW∣DERMASON)×P(LOW∣DERMASON)×P(MED∣DERMASON)×P(MED∣DERMASON)

Highest probability -> DERMASON

**HIGH HIGH HIGH HIGH:**

Probabilities for SEKER: 𝑃(𝐻𝐼𝐺𝐻∣𝑆𝐸𝐾𝐸𝑅)×𝑃(𝐻𝐼𝐺𝐻∣𝑆𝐸𝐾𝐸𝑅)×𝑃(𝐻𝐼𝐺𝐻∣𝑆𝐸𝐾𝐸𝑅)×𝑃(𝐻𝐼𝐺𝐻∣𝑆𝐸𝐾𝐸𝑅)P(HIGH∣SEKER)×P(HIGH∣SEKER)×P(HIGH∣SEKER)×P(HIGH∣SEKER)

Highest probability -> SEKER

**Decision Tree Prediction**

Using the decision tree constructed earlier:

LOW LOW MED V.HIGH:

From the decision tree: This matches BARBUNYA.

MED HIGH LOW LOW:

From the decision tree: This matches SIRA.

HIGH MED HIGH HIGH:

From the decision tree: This matches HOROZ.

V.HIGH HIGH V.HIGH HIGH:

From the decision tree: This matches CALI.

LOW LOW MED MED:

From the decision tree: This matches DERMASON.

HIGH HIGH HIGH HIGH:

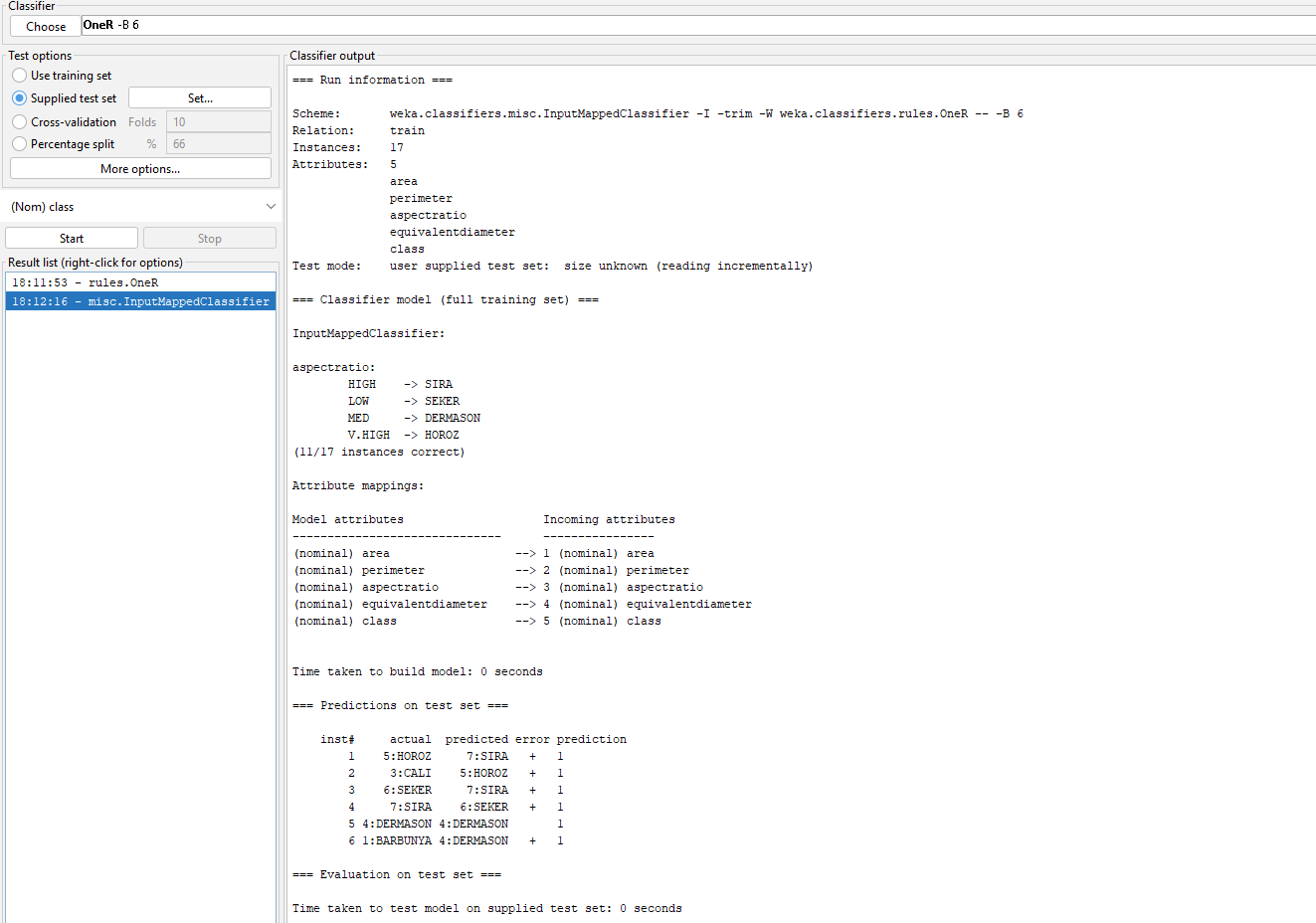
From the decision tree: This matches SEKER

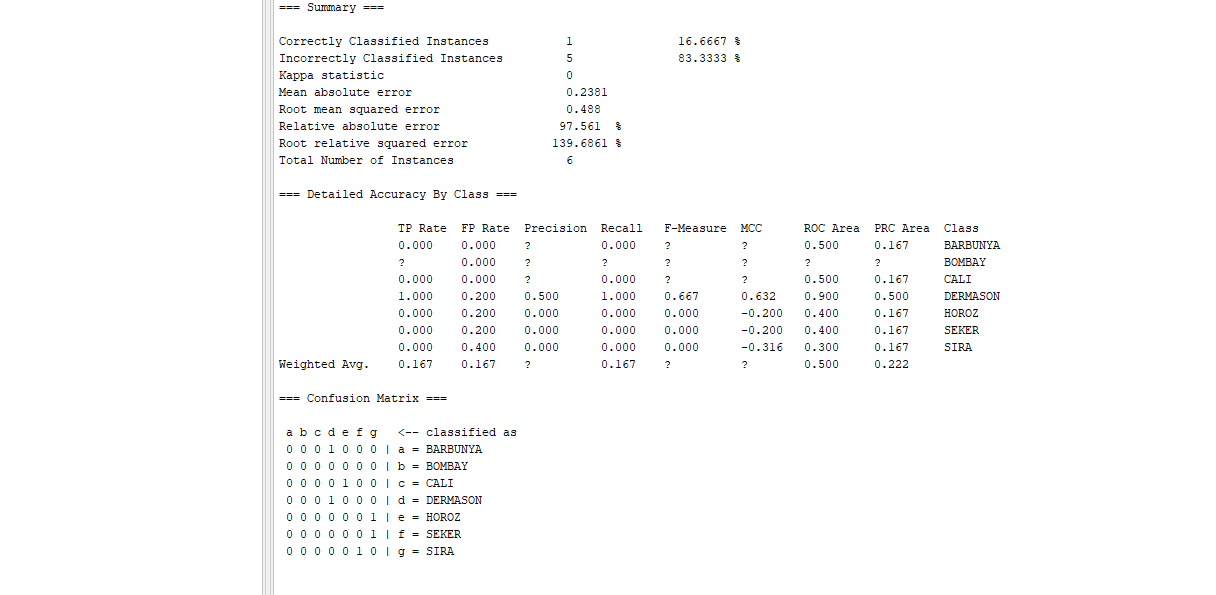
**Summary of Prediction**

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | **R1 Prediction** | **Bayes Prediction** | **Decision Tree Prediction** |
| LOW LOW MED V.HIGH | BARBUNYA | BARBUNYA | BARBUNYA |
| MED HIGH LOW LOW | SIRA | SIRA | SIRA |
| HIGH MED HIGH HIGH | HOROZ | HOROZ | HOROZ |
| V.HIGH HIGH V.HIGH HIGH | CALI | CALI | CALI |
| LOW LOW MED MED | DERMASON | DERMASON | DERMASON |
| HIGH HIGH HIGH HIGH | SEKER | SEKER | SEKER |

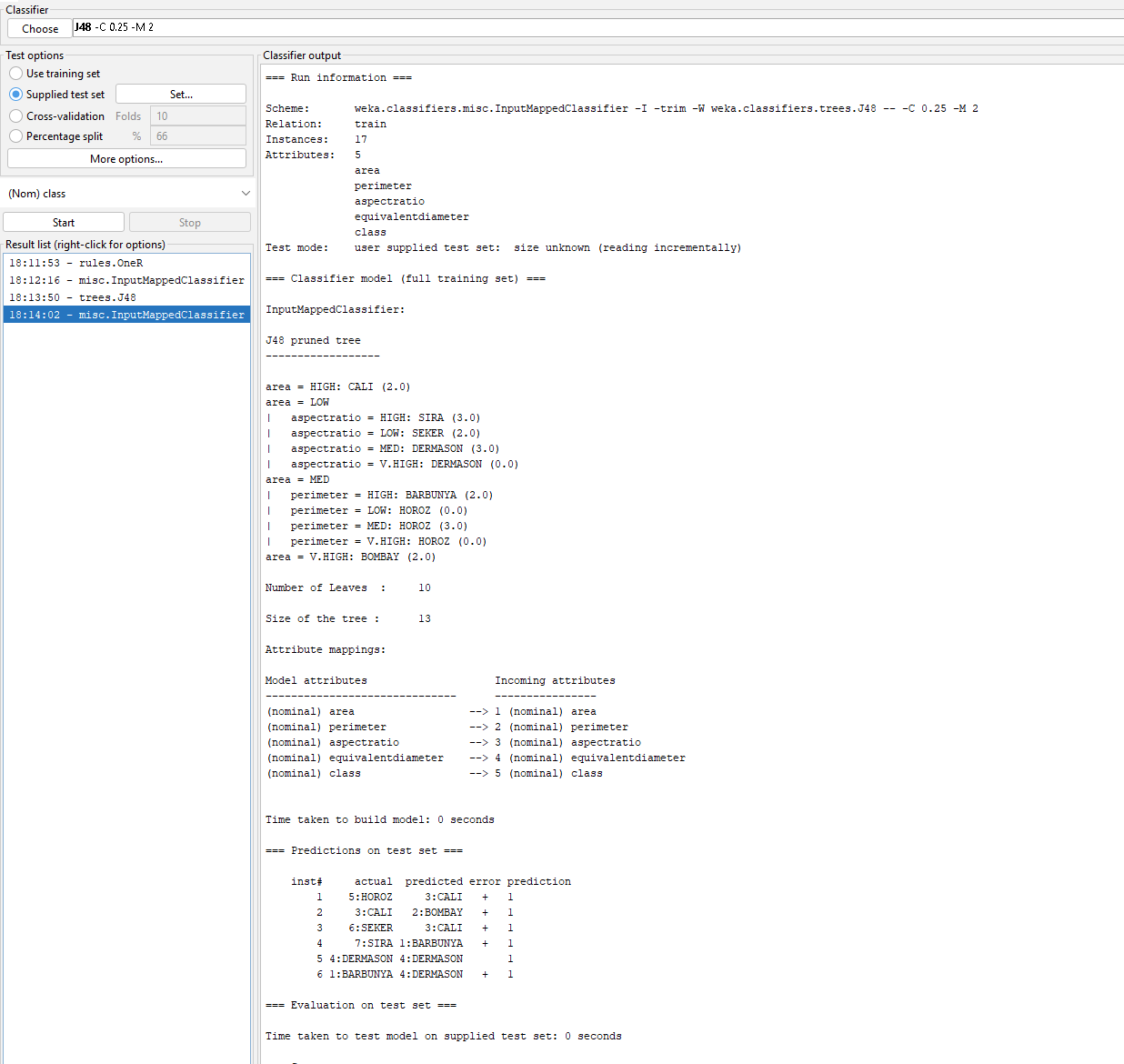
Question2: Weka

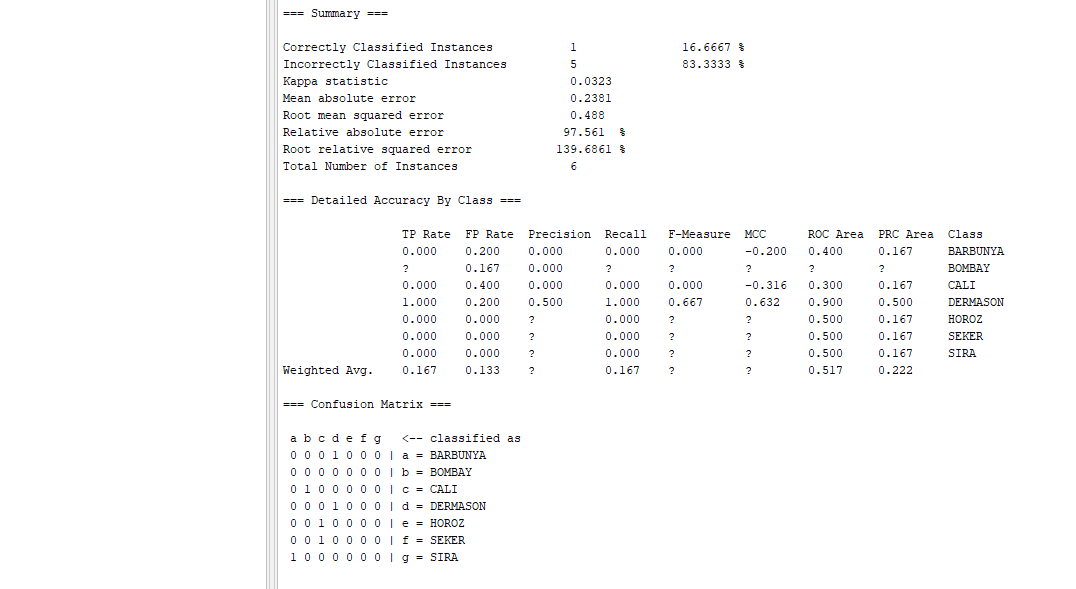
Solutions:**1R Training/Test Results**





**Training/Test Result J48**





**Test File Details**

@relation test

@attribute area {HIGH,LOW,MED,V.HIGH}

@attribute perimeter {HIGH,LOW,MED}

@attribute aspectratio {HIGH,LOW,MED,V.HIGH}

@attribute equivalentdiameter {HIGH,LOW,MED,V.HIGH}

@attribute class {BARBUNYA,CALI,DERMASON,HOROZ,SEKER,SIRA}

@data

HIGH,MED,HIGH,HIGH,HOROZ

V.HIGH,HIGH,V.HIGH,HIGH,CALI

HIGH,HIGH,HIGH,HIGH,SEKER

MED,HIGH,LOW,LOW,SIRA

LOW,LOW,MED,MED,DERMASON

LOW,LOW,MED,V.HIGH,BARBUNYA