Q1:

Generate the classification rules using OneR algorithm with the test option of 66% percentage split. Justify quantitatively how the rules were generated.



| === Run inform | nation === | | | | | | | | | |
|----------------|-------------------------------|--------------|--------------------------|-------------|-------------|-----------|---------|-----------|------------|-------|
| Scheme: | weka.classifiers.rule | s.OneR -B 6 | | | | | | | | |
| Relation: | halance-scale | | | | | | | | | |
| Instances: | 625 | | | | | | | | | |
| Attributes: | 5 | | | | | | | | | |
| NOOLIDGOOD. | left-weight | | | | | | | | | |
| | left_distance | | | | | | | | | |
| | right_weight | | | | | | | | | |
| | right_distance | | | | | | | | | |
| | class | | === Detailed A | Accuracy By | y Class ==: | = | | | | |
| Test mode. | class solit 66 0% train ra | mainder test | | _ | _ | | | | | |
| iest mode. | spiic 00.0% crain, re | mainder test | | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
| Clossifion | r model (full training | act \ | | 0.505 | 0.278 | 0.605 | 0.505 | 0.551 | 0.613 | L |
| === CIASSILIE | r moder (rurr craining | sec) === | | 0 | 0 | 0 | 0 | 0 | 0.5 | в |
| 1.5.5 | | | | 0.76 | 0.5 | 0.557 | 0.76 | 0.643 | 0.63 | R |
| lert-weight: | | | Weighted Avg. | 0.575 | 0.354 | 0.529 | 0.575 | 0.543 | 0.611 | |
| < 2.5 | -> R | | | | | | | | | |
| >= 2.3 | 5 -> Г | | === Confusion Matrix === | | | | | | | |
| (397/625 insta | ances correct) | | | | | | | | | |
| | | | abc <- | classif: | ied as | | | | | |
| | | | 49 0 48 a | a = L | | | | | | |
| Time taken to | build model: 0 second | 3 | 9010 ł | o = B | | | | | | |
| | | | 23 0 73 c | c = R | | | | | | |
| === Evaluation | n on test split === | | | | | | (falla) | | | |
| === Summary == | == | | | | | | (τοπον | wing the | iert rigui | e) |
| Correctly Clas | ssified Instances | 122 | 57.5472 % | | | | | | | |
| Incorrectly Ci | lassified Instances | 90 | 42.4528 % | | | | | | | |
| Kappa statisti | ic | 0.2216 | | | | | | | | |
| Mean absolute | error | 0.283 | | | | | | | | |
| Root mean squa | ared error | 0.532 | | | | | | | | |
| Relative absol | lute error | 74.2106 % | | | | | | | | |
| Root relative | squared error | 121.2078 % | | | | | | | | |
| Total Number o | of Instances | 212 | | | | | | | | |

The procedure of the quantitative justification (referring Figure 4.1):

| 🖻 Weka Explorer | | | | | |
|---|---------|--|-----------------------------------|--|--|
| Preprocess Classify Cluster Associate Select attributes Visualize | | | | | |
| Open file Open URL Open DB | Gener | ate Undo | Edit Save | | |
| Filter | | | | | |
| 🔲 wexa | | | Apply | | |
| AllFilter | | CSelected attribute | | | |
| MultiFilter supervised | | Name: left-weight Missing: 0 (0%) Distinct: | Type: Numeric 5 Unique: 0 (0%) | | |
| | | Statistic | Value | | |
| AddClassification | | Minimum | 1 | | |
| ClassOrder | Pattern | Maximum | 5 | | |
| Discretize | | Mean | 3 | | |
| NominalToBinary | | StdDev | 1.415 | | |
| PLSFilter | | | | | |
| 🗄 💼 instance | | | | | |
| 🖻 💼 🛅 unsupervised | | <u>[</u> | | | |
| 🗈 🛅 attribute | | Class: class (Nom) | Visualize All | | |
| | | L | | | |
| NonSparse I oSparse | | 125 125 125 | 125 125 | | |
| | | 120 120 120 | 120 120 | | |
| RemoveFolds | | | | | |
| RemoveFrequentValues | | | | | |
| RemoveMisclassified | | | | | |
| RemovePercentage | | | | | |
| RemoveRange | | | | | |
| RemoveWithValues | | 0 | 0 | | |
| Filter Remove filter Close | | r1 | 3 5 | | |

| 🎔 Weka Explorer | | | | _ 🗆 🗙 |
|---|---|--|----------------|--------------------------------|
| Preprocess Classify Cluster Asso | ciate Select attributes Visualize | | | |
| Open file Open URI | Open DB | Generate Undo | Edit | Save |
| Filter Choose Discretize -V -R last | | | | Apply |
| Current relation Relation: balance-scale Instances: 625 | Attributes: 5 | Selected attribute Name: left-weight Missing: 0 (0%) | Distinct: 5 Ur | Type: Numeric nique: 0 (0%) |
| Attributes weka.gui.GenericObjectEditor All None No. Name 1 left-weight 2 left-distance 2 left-distance 2 left-distance 2 left-distance 2 left-distance 2 left-distance | | | | |
| 4 right-distance 5 class | attributeIndices last | | ~ | Visualize All |
| | makeBinary False useBetterEncoding False useKononenko False | | | |
| R | Open S | ave OK | Cancel | 5 |

| 📚 Weka Explorer | |
|--|--|
| Preprocess Classify Cluster Associate Select attributes Visualize | |
| Open file Open URL Open DB Gene | erate Undo Edit Save |
| Filter Choose Discretize -V -R last | Apply |
| Current relation | Selected attribute |
| Relation: balance-scale-weka.filters.supervised.attribute.Discretize-V Instances: 625 Attributes: 5 | Name: left-weight Type: Nominal Missing: 0 (0%) Distinct: 2 Unique: 0 (0%) |
| Attributes | No. Label Count |
| All None Invert Pattern | 1 '(-inf-2.5]' 250 2 '(2.5-inf)' 375 |
| No. Name | |
| I left-weight 2 left-distance 3 right-weight | |
| 4 right-distance | Class: class (Nom) Visualize All |
| | 375 |
| | 250 |
| | |
| | |
| Remove | |

| 💙 Weka Explorer | | | | | | |
|--|--|--------------|--|--|--|--|
| Preprocess Classify Cluster Associate Select a | attributes Visualize | | | | | |
| Open file Open URL | Open file Open URL Open DB Generate Undo Edit Save | | | | | |
| Filter Choose RemoveWithValues -5 0.0 -C last -L first -V Apply | | | | | | |
| Current relation Relation: balance-scale-weka filters supervised | Selected attribute | inal | | | | |
| Instances: 625 Attribute | 📚 weka.gui.GenericObjectEditor 🛛 🔀 |)) | | | | |
| Attributes weka.filters.unsupervised.instance.RemoveWithValues All None Invert About | | | | | | |
| No. Name 1 left-weight | Filters instances according to the value of an attribute. More Capabilities | | | | | |
| 2 left-distance 3 right-weight 4 right-distance | attributeIndex last | isualize All | | | | |
| Class | invertSelection True | | | | | |
| | matchMissingValues False | | | | | |
| | modifyHeader False | | | | | |
| | nominalIndices first | | | | | |
| | splitPoint 0.0 | | | | | |
| Remove | Open Save OK Cancel | | | | | |

| 📚 Weka Explorer | | | | |
|--|---|--|--|--|
| Preprocess Classify Cluster Associate Select attributes Visualize | | | | |
| Open file Open URL Open DB Generate Undo Edit Save | | | | |
| Filter Choose RemoveWithValues -5 0.0 -C last -L first -V | Apply | | | |
| Current relation Relation: balance-scale-weka.filters.supervised.attribute.Discretize-V Instances: 288 Attributes: 5 | Selected attribute Name: class Type: Nominal Missing: 0 (0%) Distinct: 1 Unique: 0 (0%) | | | |
| Attributes All None Invert Pattern | No. Label | | | |
| No. Name 1 left-weight 2 left-distance 3 right-weight | | | | |
| 4 right-distance | Class: class (Nom) Visualize All | | | |
| | 288 | | | |
| Remove | 0 0 | | | |

| 📚 Weka Explorer | |
|--|---|
| Preprocess Classify Cluster Associate Select attributes Visualize | |
| Open file Open URL Open DB Gene | erate Undo Edit Save |
| Filter Choose RemoveWithValues -5 0.0 -C last -L first -V | Apply |
| Current relation Relation: balance-scale-weka.filters.supervised.attribute.Discretize-V Instances: 288 Attributes: 5 | Selected attribute Name: left-weight Type: Nominal Missing: 0 (0%) Distinct: 2 Unique: 0 (0%) |
| Attributes All None Invert Pattern | No. Label Count 1 '(-inf-2.5]' 60 2 '(2.5-inf)' 228 |
| No. Name 1 left-weight 2 left-distance 3 right-weight | |
| 4 right-distance 5 class | Class: class (Nom) Visualize All |
| | 228 |
| Remove | 60 |

You can repeat the same procedure to generate the other statistics.

| attribute | category | num | L | В | R |
|----------------|-----------|-----|-------------|-----------|-------------|
| left-weight | (-∞, 2.5] | 250 | 60 (24%) | 21 (8.4%) | 169 (67.6%) |
| | (2.5, ∞) | 375 | 228 (60.8%) | 28 (7.5%) | 119 (31.7%) |
| left-distance | (-∞, 2.5] | 250 | 60 (24%) | 21 (8.4%) | 169 (67.6%) |
| | (2.5, ∞) | 375 | 228 (60.8%) | 28 (7.5%) | 119 (31.7%) |
| right-weight | (-∞, 2.5] | 250 | 169 (67.6%) | 21 (8.4%) | 60 (24%) |
| | (2.5, ∞) | 375 | 119 (31.7%) | 28 (7.5%) | 228 (60.8%) |
| right-distance | (-∞, 2.5] | 250 | 169 (67.6%) | 21 (8.4%) | 60 (24%) |
| | (2.5, ∞) | 375 | 119 (31.7%) | 28 (7.5%) | 228 (60.8%) |

| 1. left-weight | (-∞, 2.5] => R | (2.5, ∞) => L |
|-------------------|----------------|---------------|
| 2. left-distance | (-∞, 2.5] => R | (2.5, ∞) => L |
| 3. right-weight | (-∞, 2.5] => L | (2.5, ∞) => R |
| 4. right-distance | (-∞, 2.5] => L | (2.5, ∞) => R |

Q2:

Generate a decision tree using the J48 algorithm with the test option of 66% percentage split. Was the root attribute selected based on the notion of information gain? Justify quantitatively your answer. For this purpose, assume that each numerical attribute is discretized into two ranges, the first consists of all values less than or equal to 2 and the second consists of all the values larger than 2.

A2:

The root attribute is "*left-weight*" shown as below (partly).





| 🖻 Weka Explorer | | | | | |
|---|---|--|--|--|--|
| Preprocess Classify Cluster Associate Select attributes Visualize | | | | | |
| Attribute Evaluator | | | | | |
| Choose InfoGainAttributeEval | <u>)</u> | | | | |
| | | | | | |
| Search Method | | | | | |
| Choose Ranker -T -1.797693134862 | 3157E308 -N -1 | | | | |
| Attribute Selection Mode | Attribute selection output | | | | |
| ⊙ Use full training set | === Run information === | | | | |
| O Cross-validation Folds 10 | | | | | |
| Seed 1 | Evaluator: weka.attributeSelection.InfoGainAttributeEval | | | | |
| | Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1 | | | | |
| (Nom) class 🗸 🗸 | Instances: 625 | | | | |
| | Attributes: 5 | | | | |
| Start Stop | left-weight | | | | |
| Result list (right-click for options) | left-distance | | | | |
| 17:51:17 - Ranker + InfoGainAttributeEval | right-weight | | | | |
| 17:54:15 - Ranker + GainRatioAttributeEva | right-distance | | | | |
| 17:56:07 - Ranker + InfoGainAttributeEval | Class Evaluation mode: evaluate on all training data | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | === Attribute Selection on all input data === | | | | |
| | Search Mathad: | | | | |
| | Attribute ranking. | | | | |
| | | | | | |
| | Attribute Evaluator (supervised, Class (nominal): 5 class): | | | | |
| | Information Gain Ranking Filter | | | | |
| | Ranked attributes: | | | | |
| | 0.121 2 left-distance | | | | |
| | 0.121 l left-weight | | | | |
| | 0.121 4 right-distance | | | | |
| | U.121 3 right-weight | | | | |
| | Selected attributes: 2.1.4.3 : 4 | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

The procedure of the quantitative justification:

| 📚 Weka Explorer | | | | | |
|--|---|--|--|--|--|
| Preprocess Classify Cluster Associate Select attributes Visualize | | | | | |
| Open file Open URL Open DB Gene | Open file Open URL Open DB Generate Undo Edit Save | | | | |
| Filter Choose Discretize -V -R last | Filter Choose Discretize -V -R last Apply | | | | |
| Current relation Relation: balance-scale-weka.filters.supervised.attribute.Discretize-V Instances: 625 Attributes: 5 | Selected attribute Name: left-weight Type: Nominal Missing: 0 (0%) Distinct: 2 Unique: 0 (0%) | | | | |
| Attributes All None Invert Pattern | No. Label Count (1) '(-inf-2.5]' 250 2) '(2.5-inf)' 375 | | | | |
| No. Name 1 left-weight 2 left-distance 3 right-weight | | | | | |
| 4 right-distance 5 class | Class: class (Nom) Visualize All | | | | |
| | 375 | | | | |
| | 250 | | | | |
| Remove | | | | | |

Note that since all numeric values are integers, you will get the same statistics whether you split at 2.5 or 2.0. So, you can use the previous results of question 1.

| 🎌 Weka Explorer | |
|--|--|
| Preprocess Classify Cluster Associate Select attributes Visualize | |
| Open file Open URL Open DB Gene | rate Undo Edit Save |
| Filter | |
| Choose RemoveWithValues -5 2.1 -C first -L first-last -V | Apply |
| Current relation | Selected attribute |
| Relation: balance-scale-weka.filters.unsupervised.instance.RemoveW Instances: 250 Attributes: 5 | Name: class Type: Nominal Missing: 0 (0%) Distinct: 3 Unique: 0 (0%) |
| Attributes | No. Label Count |
| | 1 L 60 |
| All None Invert Pattern | 3 R 169 |
| No. Name | |
| 1 eft-weight | |
| 2 left-distance | |
| 4 right-distance | |
| 5 class | Class: class (Nom) |
| | (<=2) |
| | |
| | 60 |
| Remove | 21 |

info(60,21,169) = entropy(60/250, 21/250, 169/250)

 $= -(60/250)\log_2(60/250) - (21/250)\log_2(21/250) - (169/250)\log_2(169/250)$

- = 0.494 + 0.3 + 0.382
- = 1.176

| 📚 Weka Explorer | | | |
|--|--------------------------------|-------------|---------------------------------|
| Preprocess Classify Cluster Associate Select attributes Visualize | | | |
| Open file Open URL Open DB Gene | rate Undo | Edit. | Save |
| Filter Choose RemoveWithValues -5 2.1 -C first -L first-last Apply | | | |
| Current relation | Selected attribute | | |
| Relation: balance-scale-weka.filters.unsupervised.instance.RemoveW Instances: 375 Attributes: 5 | Name: class Missing: 0 (0%) | Distinct: 3 | Type: Nominal Unique: 0 (0%) |
| Attributes | No. Label | | Count |
| | 1 L 2 B | | 228 |
| | 3 R | | 119 |
| No. Name | | | |
| | | | |
| 3 right-weight | | | |
| 4 right-distance | Class: class (Nom) | | Visualize All |
| 5 Class | | | |
| | 228 | (>2) | |
| | | | 119 |
| Remove | L [228] | 28 | |

info(228,28,119) = entropy(228/375, 28/375, 119/375)

 $= -(228/375)\log_2(228/375) - (28/375)\log_2(28/375) - (119/375)\log_2(119/375)$

- = 0.437 + 0.280 + 0.525
- = 1.242

```
gain (left-weight) = info (288,49,288) - info ([60,21,169], [228,28,119])
= (0.5151 + 0.288 + 0.5151) - (250/625)^{*}1.176 - (375/625)^{*}1.242
= 1.318 - 0.4704 - 0.7452
= 0.103
```

You can repeat the same procedure to get the the information gain corresponding to each of the remaining attributes.