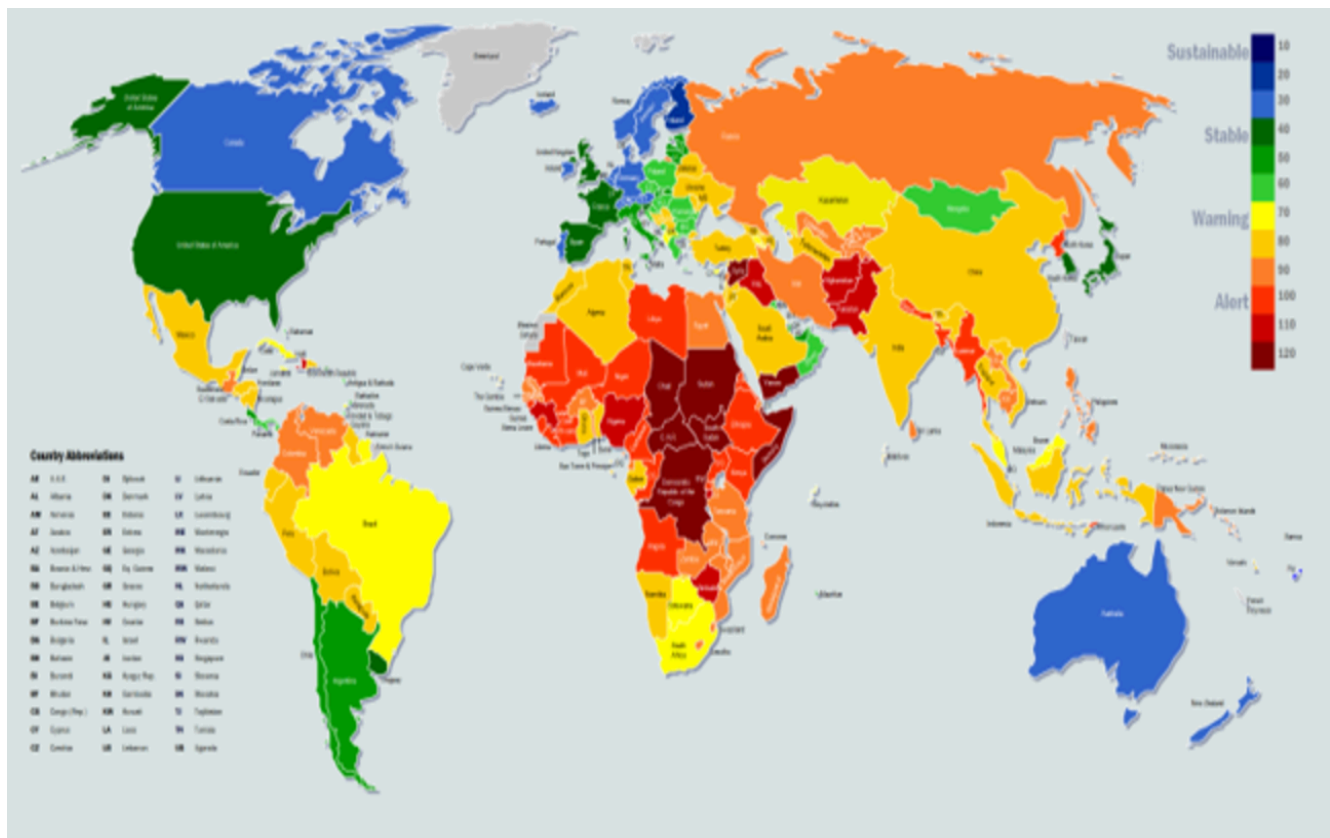


ITCS 6162: Knowledge Discovery in Databases

Project Report



Venkata Naga Akash Ungarala 800902992

Tejaswi Dumpala 800902462

Lavanya Ayila 800960067

Harini Vasudevan 800917793

Dataset

Data for the years 2011-2014 is collected from the URL: <http://fsi.fundforpeace.org/>

Tools Used

Java Compiler

WEKA 3.6 (stable version)

Data Extraction

We have used web sources like Wikipedia and annually published global reports to extract values for few new attributes to consolidate with the data set.

WEKA

Weka is an open source and popular suite of machine learning workbench that contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. The original non-Java version of Weka was a TCL/TK front-end to (mostly third-party) modeling algorithms implemented in other programming languages, plus data preprocessing utilities in C, and a Make file-based system for running machine learning experiments. This original version was primarily designed as a tool for analyzing data from agricultural domains but the more recent fully Java-based version (Weka 3), is now used in many different application areas, in particular for educational purposes and research.

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. All of Weka's techniques are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). Weka provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query. It is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using Weka. Another important area that is currently not covered by the algorithms included in the Weka distribution is sequence modeling.

Work Flow

Extraction of Structured Dataset and Pre-Processing

- ✓ We have extracted data for 178 countries from the web data. The extracted data set is cleaned and refined by replacing undefined values and records having any special characters with 0.
- ✓ The final data set contains 178 countries. The first 12 features (Demographic Pressures, Refugees and IDPs, Group Grievance, Human Flight, Uneven Development, Poverty and Economic Decline, Legitimacy of the State, Public Services, Human Rights, Security Apparatus, Factionalized Elites, External Intervention) are mandatory for the project, the other 6 features (Air Quality Index, Happiness Index, Global Warming Index, Global Peace Index, Technology Usage Index, Global Slavery Index) are the additional features added.
- ✓ Once the data is clean, it is loaded into WEKA Explorer.

Year 2011

Discretize

We have Total as the decision feature. In the data set, the total column is sorted. Since we have to split the total into 4 intervals, we can split it as the levels Alert, Warning, Stable and Sustainable. We need to choose the split to get a maximum entropy gain for the decision feature considered.

Alert 38

Warning 83

Stable 33

Sustainable 23

Entropy Gain for Total

$$E(D) = -(((38/177) * \log_4(38/177)) + ((83/177) * \log_4(83/177)) + ((33/177) * \log_4(33/177)) + ((23/177) * \log_4(23/177)))$$

$$E(D) = -(0.214689266 * (-1.109839018) + 0.468926554 * (-0.546283059) + 0.186440678 * (-1.211605715) + 0.129943503 * (-1.472021797))$$

$$E(D) = 0.911609416$$

We have calculated the initial entropy and made the split based on the entropy gain of the split. We have chosen the split which has higher Entropy Gain and decided to have the total split into 4 intervals as below

(126.6 ... +infinity)

(94.1 ... 126.6]

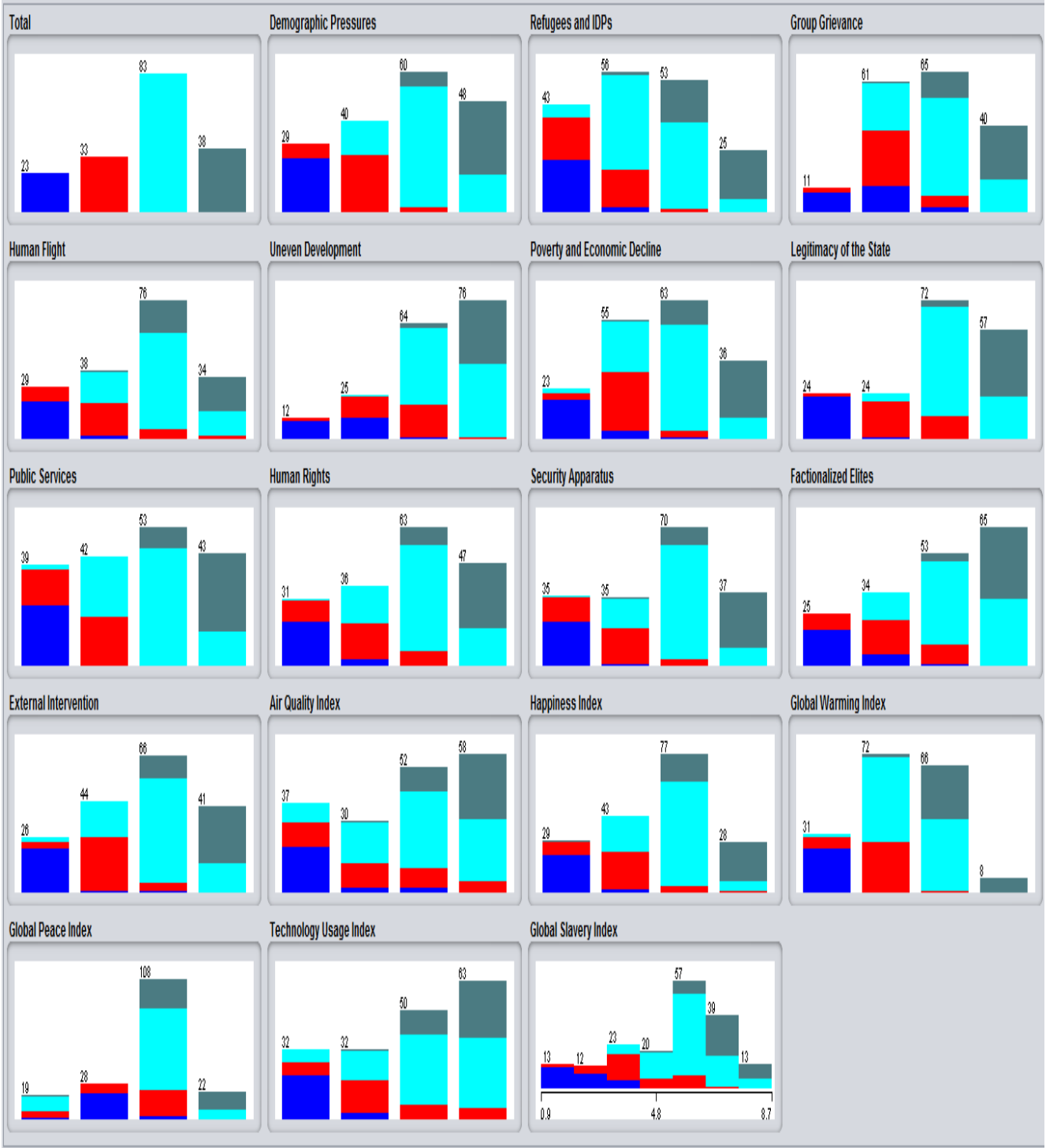
(61.6 ... 94.1]

(-infinity ... 61.6]

Classification using WEKA

Select Total as the decision feature and apply different algorithms to compare the results and choose the algorithm that gives the best number of correctly classified records.

Visualize all the features



Bayes - BayesNet Classifier

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **HoeffdingTree-L2-S1-E1.0E-7-H0.05-M0.01-G200.0-N0.0**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 18:09:51 - trees.RandomTree
- 18:10:02 - trees.RandomForest
- 18:10:08 - bayes.NaiveBayes
- 18:10:17 - bayes.BayesNet**
- 18:11:10 - trees.HoeffdingTree

Classifier output

```

Correctly Classified Instances      164      92.6554 %
Incorrectly Classified Instances    13      7.3446 %
Kappa statistic                    0.8919
Mean absolute error                 0.0373
Root mean squared error             0.161
Relative absolute error             10.8931 %
Root relative squared error         38.9584 %
Total Number of Instances          177

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      1.000    0.006    0.958     1.000    0.979     0.976    1.000     1.000     '(-inf-61.6]'
      0.848    0.014    0.933     0.848    0.889     0.866    0.996     0.982     '(61.6-94.1]'
      0.940    0.074    0.918     0.940    0.929     0.864    0.990     0.988     '(94.1-126.6]'
      0.921    0.022    0.921     0.921    0.921     0.899    0.995     0.983     '(126.6-inf)'
Weighted Avg.    0.927    0.043    0.927     0.927    0.926     0.887    0.993     0.988

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
23  0  0  0 | a = '(-inf-61.6]'
 1 28  4  0 | b = '(61.6-94.1]'
 0  2 78  3 | c = '(94.1-126.6]'
 0  0  3 35 | d = '(126.6-inf)'
  
```

Classifier Output

Correctly Classified Instances 164 92.6554 %

Incorrectly Classified Instances 13 7.3446 %

Kappa statistic 0.8919

Mean absolute error 0.0373

Root mean squared error 0.161

Relative absolute error 10.8931 %

Root relative squared error 38.9584 %

Total Number of Instances 177

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.006	0.958	1.000	0.979	0.976	1.000	1.000	'(-inf-61.6]'
0.848	0.014	0.933	0.848	0.889	0.866	0.996	0.982	'(61.6-94.1]'
0.940	0.074	0.918	0.940	0.929	0.864	0.990	0.988	'(94.1-126.6]'
0.921	0.022	0.921	0.921	0.921	0.899	0.995	0.983	'(126.6-inf)'

Confusion Matrix

	a	b	c	d	
a	23	0	0	0	a = '(-inf - 61.6]'
b	1	28	4	0	b = '(61.6 - 94.1]'
c	0	2	78	3	c = '(94.1 - 126.6]'
d	0	0	3	35	d = '(126.6 - inf)'

Bayes - Naive Bayes Classifier

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **HoeffdingTree** -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 18:09:51 - trees.RandomTree
- 18:10:02 - trees.RandomForest
- 18:10:08 - bayes.NaiveBayes
- 18:10:17 - bayes.BayesNet
- 18:11:10 - trees.HoeffdingTree

Classifier output

```

=== Summary ===
Correctly Classified Instances      168      94.9153 %
Incorrectly Classified Instances    9        5.0847 %
Kappa statistic                    0.9255
Mean absolute error                 0.036
Root mean squared error             0.1568
Relative absolute error             10.5086 %
Root relative squared error         37.9415 %
Total Number of Instances          177

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      1.000    0.006    0.958     1.000    0.979     0.976    1.000    1.000    '(-inf-61.6]'
      0.909    0.014    0.938     0.909    0.923     0.906    0.995    0.981    '(61.6-94.1]'
      0.952    0.043    0.952     0.952    0.952     0.909    0.989    0.986    '(94.1-126.6]'
      0.947    0.014    0.947     0.947    0.947     0.933    0.993    0.982    '(126.6-inf]'
Weighted Avg.  0.949    0.026    0.949     0.949    0.949     0.922    0.992    0.986

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
23  0  0  0  | a = '(-inf-61.6]'
 1 30  2  0  | b = '(61.6-94.1]'
 0  2 79  2  | c = '(94.1-126.6]'
 0  0  3 36  | d = '(126.6-inf]'

```

Classifier Output

Correctly Classified Instances	168	94.9153 %
Incorrectly Classified Instances	9	5.0847 %
Kappa statistic	0.9255	
Mean absolute error	0.036	
Root mean squared error	0.1568	
Relative absolute error	10.5086 %	
Root relative squared error	37.9415 %	
Total Number of Instances	177	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.006	0.958	1.000	0.979	0.976	1.000	1.000	'(-inf-61.6]'
0.909	0.014	0.938	0.909	0.923	0.906	0.995	0.981	'(61.6-94.1]'
0.952	0.043	0.952	0.952	0.952	0.909	0.989	0.986	'(94.1-126.6]'
0.947	0.014	0.947	0.947	0.947	0.933	0.993	0.982	'(126.6-inf)'

Confusion Matrix

	a	b	c	d	
a	23	0	0	0	a = '(-inf - 61.6]'
b	1	30	2	0	b = '(61.6 - 94.1]'
c	0	2	79	2	c = '(94.1 - 126.6]'
d	0	0	2	36	d = '(126.6 - inf)'

Trees - Random Forest Classifier

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize

Choose
HoeffdingTree -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation
☐ Percentage split

Folds: 10

More options...

(Nom) Total

Start
Stop

Result list (right-click for options)

Classifier output

```

=== Summary ===
Correctly Classified Instances      160          90.3955 %
Incorrectly Classified Instances    17           9.6045 %
Kappa statistic                    0.8589
Mean absolute error                 0.106
Root mean squared error             0.1964
Relative absolute error             30.988 %
Root relative squared error         47.5402 %
Total Number of Instances          177

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.957    0.006    0.957     0.957    0.957     0.950    0.999     0.997    '(-inf-61.6]
      0.818    0.028    0.871     0.818    0.844     0.810    0.988     0.956    '(61.6-94.1]
      0.916    0.085    0.905     0.916    0.910     0.830    0.973     0.966    '(94.1-126.6]
      0.921    0.029    0.897     0.921    0.909     0.884    0.989     0.970    '(126.6-inf)
Weighted Avg.    0.904    0.052    0.904     0.904    0.904     0.853    0.983     0.969

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
22  1  0  0  | a = '(-inf-61.6]
 1 27  5  0  | b = '(61.6-94.1]
 0  3 76  4  | c = '(94.1-126.6]
 0  0  3 35  | d = '(126.6-inf)

```

Classifier Output

Correctly Classified Instances	160	90.3955 %
Incorrectly Classified Instances	17	9.6045 %
Kappa statistic	0.8589	
Mean absolute error	0.106	
Root mean squared error	0.1964	
Relative absolute error	30.988 %	
Root relative squared error	47.5402 %	
Total Number of Instances	177	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.957	0.006	0.957	0.957	0.957	0.950	0.999	0.997	'(-inf-61.6]'
0.818	0.028	0.871	0.818	0.844	0.810	0.988	0.956	'(61.6-94.1]'
0.916	0.085	0.905	0.916	0.910	0.830	0.973	0.966	'(94.1-126.6]'
0.921	0.029	0.897	0.921	0.909	0.884	0.989	0.970	'(126.6-inf)'

Confusion Matrix

a	b	c	d	
22	1	0	0	a = '(-inf - 61.6]'
1	27	5	0	b = '(61.6 - 94.1]'
0	3	76	4	c = '(94.1 - 126.6]'
0	0	3	35	d = '(126.6 - inf)'

Trees - Random Tree Classifier

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize

Classifier

Choose **HoeffdingTree** -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds
☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 18:09:51 - trees.RandomTree
- 18:10:02 - trees.RandomForest
- 18:10:08 - bayes.NaiveBayes
- 18:10:17 - bayes.BayesNet
- 18:11:10 - trees.HoeffdingTree

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      142      80.226 %
Incorrectly Classified Instances    35      19.774 %
Kappa statistic                    0.7142
Mean absolute error                 0.098
Root mean squared error             0.3115
Relative absolute error             28.6287 %
Root relative squared error         75.387 %
Total Number of Instances          177

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
      1.000   0.019   0.885     1.000   0.939   0.931   0.990   0.885   '(-inf-61.6]
      0.697   0.042   0.793     0.697   0.742   0.690   0.842   0.628   '(61.6-94.1]
      0.771   0.138   0.831     0.771   0.800   0.637   0.819   0.757   '(94.1-126.6]
      0.842   0.094   0.711     0.842   0.771   0.706   0.873   0.627   '(126.6-inf)
Weighted Avg.   0.802   0.095   0.805     0.802   0.801   0.700   0.857   0.722

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
23  0  0  0 | a = '(-inf-61.6]
 3 23  7  0 | b = '(61.6-94.1]
 0  6 64 13 | c = '(94.1-126.6]
 0  0  6 32 | d = '(126.6-inf)

```

Classifier Output

Correctly Classified Instances	142	80.226 %
Incorrectly Classified Instances	35	19.774 %
Kappa statistic	0.7142	
Mean absolute error	0.098	
Root mean squared error	0.3115	
Relative absolute error	28.6287 %	
Root relative squared error	75.387 %	
Total Number of Instances	177	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.019	0.885	1.000	0.939	0.931	0.990	0.885	'(-inf-61.6]'
0.697	0.042	0.793	0.697	0.742	0.690	0.842	0.628	'(61.6-94.1]'
0.771	0.138	0.831	0.771	0.800	0.637	0.819	0.757	'(94.1-126.6]'
0.842	0.094	0.711	0.842	0.771	0.706	0.873	0.627	'(126.6-inf)'

Confusion Matrix

	a	b	c	d	
23	0	0	0		a = '(-inf - 61.6]'
3	23	7	0		b = '(61.6 - 94.1]'
0	6	64	13		c = '(94.1 - 126.6]'
0	0	6	32		d = '(126.6 - inf)'

Trees – Hoeffding Tree Classifier

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize

Choose
HoeffdingTree -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Total
Start Stop

Result list (right-click for options)
18:09:51 - trees.RandomTree
18:10:02 - trees.RandomForest
18:10:08 - bayes.NaiveBayes
18:10:17 - bayes.BayesNet
18:11:10 - trees.HoeffdingTree

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      168      94.9153 %
Incorrectly Classified Instances     9       5.0847 %
Kappa statistic                    0.9255
Mean absolute error                 0.0359
Root mean squared error             0.1569
Relative absolute error             10.4982 %
Root relative squared error         37.966 %
Total Number of Instances          177

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
1.000   0.006   0.958   1.000   0.979   0.976   1.000   1.000   '(-inf-61.6]
0.909   0.014   0.938   0.909   0.923   0.906   0.995   0.981   '(61.6-94.1]
0.952   0.043   0.952   0.952   0.952   0.909   0.989   0.986   '(94.1-126.6]
0.947   0.014   0.947   0.947   0.947   0.933   0.993   0.982   '(126.6-inf)
Weighted Avg.   0.949   0.026   0.949   0.949   0.949   0.922   0.992   0.986

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
23  0  0  0 | a = '(-inf-61.6]
 1 30  2  0 | b = '(61.6-94.1]
 0  2 79  2 | c = '(94.1-126.6]
 0  0  2 36 | d = '(126.6-inf)

```

Classifier Output

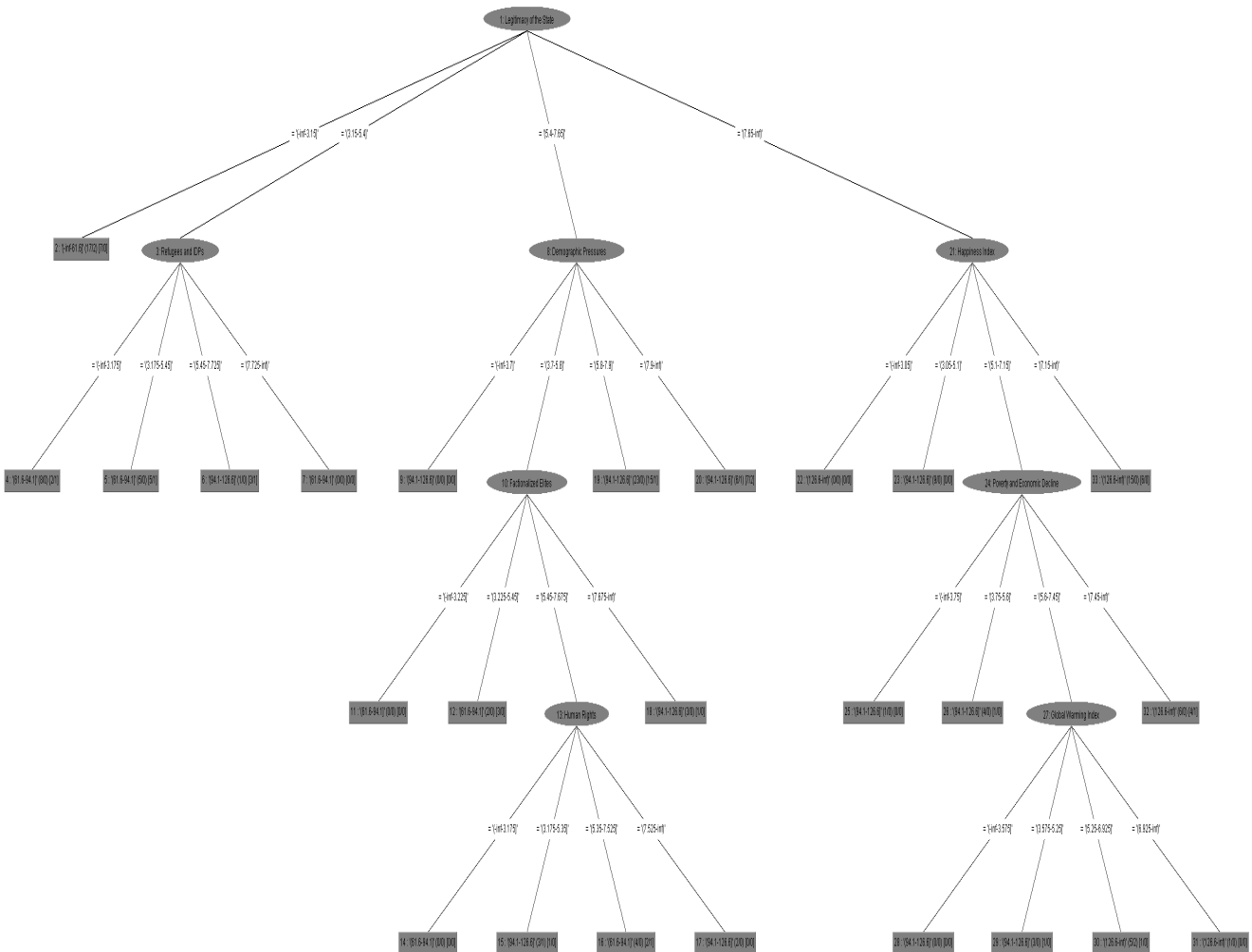
Correctly Classified Instances	168	94.9153 %
Incorrectly Classified Instances	9	5.0847 %
Kappa statistic	0.9255	
Mean absolute error	0.0359	
Root mean squared error	0.1569	
Relative absolute error	10.4982 %	
Root relative squared error	37.966 %	
Total Number of Instances	177	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.006	0.958	1.000	0.979	0.976	1.000	1.000	'(-inf-61.6]'
0.909	0.014	0.938	0.909	0.923	0.906	0.995	0.981	'(61.6-94.1]'
0.952	0.043	0.952	0.952	0.952	0.909	0.989	0.986	'(94.1-126.6]'
0.947	0.014	0.947	0.947	0.947	0.933	0.993	0.982	'(126.6-inf)'

Confusion Matrix

	a	b	c	d	
a	23	0	0	0	a = '(-inf - 61.6]'
b	1	30	2	0	b = '(61.6 - 94.1]'
c	0	2	79	2	c = '(94.1 - 126.6]'
d	0	0	2	36	d = '(126.6 - inf)'

Tree Visualizer



Comparison of Algorithms and Result

Since the percentage of correctly classified instances for Bayes Net, Naïve Bayes, Random Forest, Random Tree and Hoeffding Tree algorithms is 92.6554, 94.9153, 90.3955, 80.226 and 94.9153 respectively. This clearly shows that Naïve Bayes and Hoeffding Tree algorithms are best used to classify as it has high accuracy when compared with other algorithms.

Determining Action Rules

We have determined the action rules by using Weka by selecting the Rules-PART filter as the classifier and run for the data with total as the decision attribute. The rules are logged in a word document and the result screenshot is shown below.

The screenshot shows the Weka software interface with the 'Classifier' tab selected. The 'PART-M 2-C 0.25-Q 1-output-debug-info' classifier is chosen. The 'Test options' section shows 'Cross-validation Folds 10' selected. The 'Classifier output' section displays a list of PART decision rules for the 'Total' attribute. The rules are as follows:

```
PART decision list
-----

Legitimacy of the State = '(-inf-3.15]' AND
Demographic Pressures = '(-inf-3.7]': '(-inf-61.6]' (22.0)

Demographic Pressures = '(5.8-7.9]' AND
Legitimacy of the State = '(5.4-7.65]': '(94.1-126.6]' (38.0/1.0)

Uneven Development = '(7.225-inf)' AND
Security Apparatus = '(7.75-inf)': '(126.6-inf)' (30.0/2.0)

Global Warming Index = '(3.575-5.25]' AND
Group Grievance = '(3.225-5.45]' AND
Poverty and Economic Decline = '(3.75-5.6]': '(61.6-94.1]' (17.0)

Uneven Development = '(3.275-5.25]': '(61.6-94.1]' (10.0/1.0)

Human Flight = '(5.2-7.25]' AND
Global Warming Index = '(3.575-5.25]': '(94.1-126.6]' (14.0)

Refugees and IDPs = '(3.175-5.45]' AND
Security Apparatus = '(5.5-7.75]': '(94.1-126.6]' (12.0)

Happiness Index = '(5.1-7.15]' AND
External Intervention = '(5.55-7.775]': '(94.1-126.6]' (14.0/2.0)

Demographic Pressures = '(7.9-inf)' AND
```

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **PART - M 2-C 0.25-Q 1 -output-debug-info**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

(Num) Global Slavery Index

Result list (right-click for options)

- 14:16:03 - bayes.BayesNet
- 14:44:09 - rules.PART
- 14:47:26 - rules.PART

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      132           74.5763 %
Incorrectly Classified Instances    45           25.4237 %
Kappa statistic                    0.6232
Mean absolute error                 0.1346
Root mean squared error             0.3385
Relative absolute error             39.3266 %
Root relative squared error         81.9062 %
Total Number of Instances          177

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.957   0.019   0.880   0.957   0.917   0.905   0.970   0.868   '(-inf-61.6]'
0.515   0.049   0.708   0.515   0.596   0.531   0.741   0.543   '(61.6-94.1]'
0.795   0.234   0.750   0.795   0.772   0.560   0.829   0.751   '(94.1-126.6]'
0.711   0.094   0.675   0.711   0.692   0.606   0.887   0.640   '(126.6-inf)'
Weighted Avg.  0.746   0.141   0.743   0.746   0.741   0.609   0.843   0.704

=== Confusion Matrix ===

 a b c d <-- classified as
22 1 0 0 | a = '(-inf-61.6]'
 3 17 11 2 | b = '(61.6-94.1]'
 0 6 66 11 | c = '(94.1-126.6]'
 0 0 11 27 | d = '(126.6-inf)'

```

Year 2012

Discretize

We have Total as the decision feature. In the data set, the total column is sorted. Since we have to split the total into 4 intervals, we can split it as the levels Alert, Warning, Stable and Sustainable. We need to choose the split to get a maximum entropy gain for the decision feature considered.

Alert 33

Warning 80

Stable 41

Sustainable 24

Entropy Gain for Total

$$E(D) = -((33/178) * \log_4(33/178)) + ((80/178) * \log_4(80/178)) + ((41/178) * \log_4(41/178)) + ((24/178) * \log_4(24/178))$$

$$E(D) = -(0.185393258 * (-1.215669656)) + 0.449438202 * (-0.576902668) + 0.230337079 * (-1.059090713) + 0.134831461 * (-1.445385465)$$

$$E(D) = 0.923490351$$

We have calculated the initial entropy and made the split based on the entropy gain of the split. We have chosen the split which has higher Entropy Gain and decided to have the total split into 4 intervals as below

(131.775 ... +infinity)

(97.65 ... 131.775]

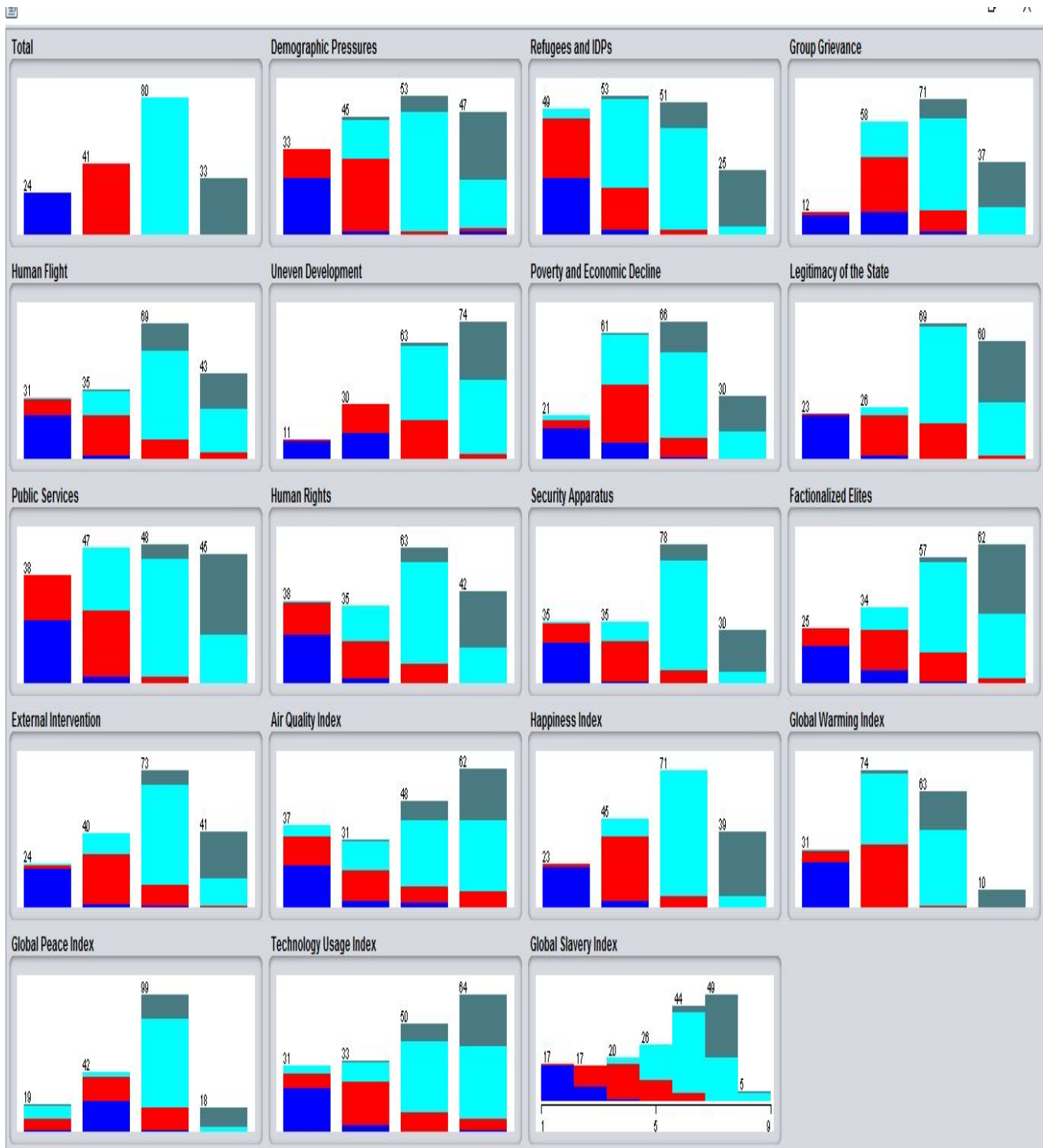
(63.525 ... 97.65]

(-infinity ... 63.525]

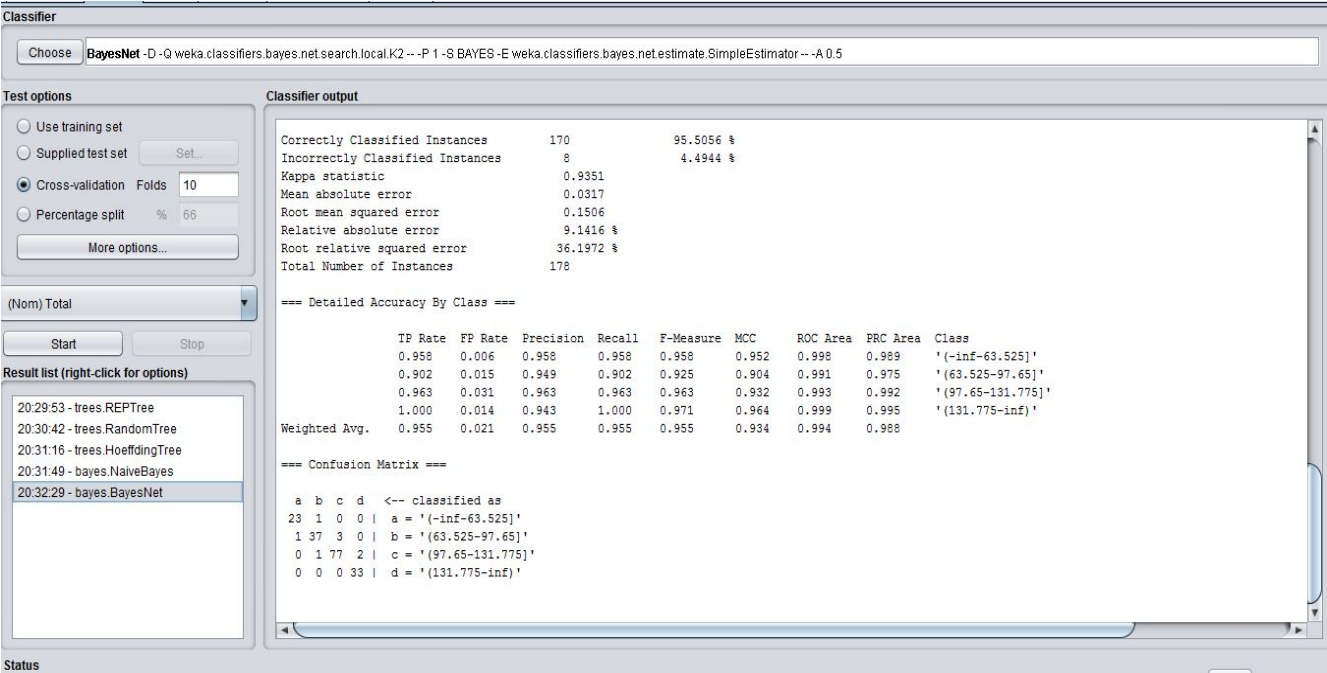
Classification using WEKA

Select Total as the decision feature and apply different algorithms to compare the results and choose the algorithm that gives the best number of correctly classified records.

Visualize all the features



Bayes - BayesNet Classifier



Classifier Output

Correctly Classified Instances	170	95.5056 %
Incorrectly Classified Instances	8	4.4944 %
Kappa statistic	0.9351	
Mean absolute error	0.0317	
Root mean squared error	0.1506	
Relative absolute error	9.1416 %	
Root relative squared error	36.1972 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.958	0.006	0.958	0.958	0.958	0.952	0.998	0.989	'(-inf-63.525]'
0.902	0.015	0.949	0.902	0.925	0.904	0.991	0.975	'(63.525-97.65]'
0.963	0.031	0.963	0.963	0.963	0.932	0.993	0.992	'(97.65-131.775]'
1.000	0.014	0.943	1.000	0.971	0.964	0.999	0.995	'(131.775-inf)'

Confusion Matrix

a	b	c	d	
23	1	0	0	a = '(-inf - 63.525]'
1	37	3	0	b = '(63.525 - 97.65]'
0	1	77	2	c = '(97.65 - 131.775]'
0	0	0	33	d = '(131.775 - inf)'

Bayes - Naive Bayes Classifier

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 20:29:53 - trees.REPTree
- 20:30:42 - trees.RandomTree
- 20:31:16 - trees.HoeffdingTree
- 20:31:49 - bayes.NaiveBayes

Classifier output

```

=== Summary ===
Correctly Classified Instances      167      93.8202 %
Incorrectly Classified Instances    11       6.1798 %
Kappa statistic                    0.9113
Mean absolute error                 0.0322
Root mean squared error             0.1471
Relative absolute error             9.2759 %
Root relative squared error        35.341 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.958  0.006  0.958  0.958  0.958  0.952  0.999  0.991  '(-inf-63.525]'
0.902  0.029  0.902  0.902  0.902  0.873  0.992  0.975  '(63.525-97.65]'
0.925  0.031  0.961  0.925  0.943  0.898  0.993  0.992  '(97.65-131.775]'
1.000  0.021  0.917  1.000  0.957  0.947  0.999  0.995  '(131.775-inf)'
Weighted Avg.  0.938  0.025  0.939  0.938  0.938  0.909  0.995  0.989

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
23  1  0  0  a = '(-inf-63.525]'
 1 37  3  0  b = '(63.525-97.65]'
 0  3 74  3  c = '(97.65-131.775]'
 0  0  0 33  d = '(131.775-inf)'

```

Classifier Output

Correctly Classified Instances	167	93.8202 %
Incorrectly Classified Instances	11	6.1798 %
Kappa statistic	0.9113	
Mean absolute error	0.0322	
Root mean squared error	0.1471	
Relative absolute error	9.2759 %	
Root relative squared error	35.341 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.958	0.006	0.958	0.958	0.958	0.952	0.999	0.991	'(-inf-63.525]'
0.902	0.029	0.902	0.902	0.902	0.873	0.992	0.975	'(63.525-97.65]'
0.925	0.031	0.961	0.925	0.943	0.898	0.993	0.992	'(97.65-131.775]'
1.000	0.021	0.917	1.000	0.957	0.947	0.999	0.995	'(131.775-inf)'

Confusion Matrix

a	b	c	d	
23	1	0	0	a = '(-inf - 63.525]'
1	37	3	0	b = '(63.525 - 97.65]'
0	3	74	3	c = '(97.65 - 131.775]'
0	0	0	33	d = '(131.775 - inf)'

Trees – Random Tree Classifier

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **RandomTree-K 0-M 1.0-V 0.001-S 1**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 20:29:53 - trees.REPTree
- 20:30:42 - trees.RandomTree

Classifier output

```

=== Summary ===
Correctly Classified Instances      145           81.4607 %
Incorrectly Classified Instances    33           18.5393 %
Kappa statistic                    0.7329
Mean absolute error                 0.0912
Root mean squared error             0.2996
Relative absolute error             26.2761 %
Root relative squared error         71.9905 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.792    0.032    0.792    0.792    0.792    0.759    0.881    0.682    '(-inf-63.525]
0.707    0.080    0.725    0.707    0.716    0.633    0.839    0.618    '(63.525-97.65]
0.850    0.112    0.861    0.850    0.855    0.739    0.879    0.814    '(97.65-131.775]
0.879    0.041    0.829    0.879    0.853    0.819    0.919    0.751    '(131.775-inf)'
Weighted Avg.    0.815    0.081    0.814    0.815    0.814    0.732    0.877    0.739

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
19  5  0  0  | a = '(-inf-63.525]'
 5 29  7  0  | b = '(63.525-97.65]'
 0  6 68  6  | c = '(97.65-131.775]'
 0  0  4 29  | d = '(131.775-inf)'

```

Classifier Output

Correctly Classified Instances	145	81.4607 %
Incorrectly Classified Instances	33	18.5393 %
Kappa statistic	0.7329	
Mean absolute error	0.0912	
Root mean squared error	0.2996	
Relative absolute error	26.2761 %	
Root relative squared error	71.9905 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.792	0.032	0.792	0.792	0.792	0.759	0.881	0.682	'(-inf-63.525]'
0.707	0.080	0.725	0.707	0.716	0.633	0.839	0.618	'(63.525-97.65]'
0.850	0.112	0.861	0.850	0.855	0.739	0.879	0.814	'(97.65-131.775]'
0.879	0.041	0.829	0.879	0.853	0.819	0.919	0.751	'(131.775-inf)'

Confusion Matrix

a	b	c	d	
19	5	0	0	a = '(-inf - 63.525]'
5	29	7	0	b = '(63.525 - 97.65]'
0	6	68	6	c = '(97.65 - 131.775]'
0	0	4	29	d = '(131.775 - inf)'

Trees – REP Tree Classifier

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier

Choose **REPtree-M 2-V 0.001-N 3-S 1-L-1-I 0.0**

Test options

☐ Use training set

☐ Supplied test set **Set...**

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

More options...

(Nom) Total

Start **Stop**

Result list (right-click for options)

20:29:53 - trees.REPtree

Classifier output

```

=== Summary ===
Correctly Classified Instances      151           84.8315 %
Incorrectly Classified Instances    27           15.1685 %
Kappa statistic                    0.785
Mean absolute error                 0.1152
Root mean squared error             0.2589
Relative absolute error             33.1968 %
Root relative squared error        62.2107 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.875   0.026   0.840   0.875   0.857   0.835   0.959   0.690   '(-inf-63.525]'
0.780   0.080   0.744   0.780   0.762   0.689   0.847   0.723   '(63.525-97.65]'
0.813   0.061   0.915   0.813   0.861   0.763   0.886   0.849   '(97.65-131.775]'
1.000   0.041   0.846   1.000   0.917   0.901   0.972   0.807   '(131.775-inf)'
Weighted Avg.   0.848   0.057   0.853   0.848   0.848   0.781   0.903   0.791

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
21  3  0  0 | a = '(-inf-63.525]'
 3 32  6  0 | b = '(63.525-97.65]'
 1  8 65  6 | c = '(97.65-131.775]'
 0  0  0 33 | d = '(131.775-inf)'

```

Status

Classifier Output

Correctly Classified Instances	151	84.8315 %
Incorrectly Classified Instances	27	15.1685 %
Kappa statistic	0.785	
Mean absolute error	0.1152	
Root mean squared error	0.2589	
Relative absolute error	33.1968 %	
Root relative squared error	62.2107 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.875	0.026	0.840	0.875	0.857	0.835	0.959	0.690	'(-inf-63.525]'
0.780	0.080	0.744	0.780	0.762	0.689	0.847	0.723	'(63.525-97.65]'
0.813	0.061	0.915	0.813	0.861	0.763	0.886	0.849	'(97.65-131.775]'
1.000	0.041	0.846	1.000	0.917	0.901	0.972	0.807	'(131.775-inf)'

Confusion Matrix

a	b	c	d	
21	3	0	0	a = '(-inf - 63.525]'
3	32	6	0	b = '(63.525 - 97.65]'
1	8	65	6	c = '(97.65 - 131.775]'
0	0	0	33	d = '(131.775 - inf)'

Trees – Hoeffding Tree Classifier

Classifier

Choose **HoeffdingTree** -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Total

Result list (right-click for options)

20:29:53 - trees.REPtree

20:30:42 - trees.RandomTree

20:31:16 - trees.HoeffdingTree

Classifier output

```

Summary
Correctly Classified Instances      168      94.382 %
Incorrectly Classified Instances    10       5.618 %
Kappa statistic                    0.9194
Mean absolute error                 0.0321
Root mean squared error             0.1466
Relative absolute error              9.2407 %
Root relative squared error         35.2148 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      1.000    0.006    0.960     1.000    0.980     0.977    0.999    0.993    '(-inf-63.525]'
      0.902    0.022    0.925     0.902    0.914     0.888    0.992    0.976    '(63.525-97.65]'
      0.925    0.031    0.961     0.925    0.943     0.898    0.993    0.992    '(97.65-131.775]'
      1.000    0.021    0.917     1.000    0.957     0.947    0.999    0.995    '(131.775-inf)'
Weighted Avg.    0.944    0.024    0.944     0.944    0.944     0.916    0.995    0.989

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
24  0  0  0  | a = '(-inf-63.525]'
 1 37  3  0  | b = '(63.525-97.65]'
 0  3 74  3  | c = '(97.65-131.775]'
 0  0  0 33  | d = '(131.775-inf)'

```

Classifier Output

Correctly Classified Instances	168	94.382 %
Incorrectly Classified Instances	10	5.618 %
Kappa statistic	0.9194	
Mean absolute error	0.0321	
Root mean squared error	0.1466	
Relative absolute error	9.2407 %	
Root relative squared error	35.2148 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.006	0.960	1.000	0.980	0.977	0.999	0.993	'(-inf-63.525]'
0.902	0.022	0.925	0.902	0.914	0.888	0.992	0.976	'(63.525-97.65]'
0.925	0.031	0.961	0.925	0.943	0.898	0.993	0.992	'(97.65-131.775]'
1.000	0.021	0.917	1.000	0.957	0.947	0.999	0.995	'(131.775-inf)'

Confusion Matrix

a	b	c	d	
24	0	0	0	a = '(-inf - 63.525]'
1	37	3	0	b = '(63.525 - 97.65]'
0	3	74	3	c = '(97.65 - 131.775]'
0	0	0	33	d = '(131.775 - inf)'

Comparison of Algorithms and Result

Since the percentage of correctly classified instances for Bayes Net, Naïve Bayes, Random Tree, REP Tree and Hoeffding Tree algorithms is 95.5056, 93.8202, 81.4607, 84.8315 and 94.382 respectively. This clearly shows that Bayes Net algorithm is best used to classify as it has high accuracy when compared with other algorithms.

Determining Action Rules

We have determined the action rules by using Weka by selecting the Rules-PART filter as the classifier and run for the data with total as the decision attribute. The rules are logged in a word document and the result screenshot is shown below.

The screenshot displays the Weka software interface, specifically the Classifier tab. The 'PART-M 2-C 0.25-Q 1-output-debug-info' classifier is selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the 'PART decision list' with 9 rules. The 'Result list (right-click for options)' pane shows a log of the execution, with the last entry being '14:47:26 - rules.PART'.

Classifier

Choose **PART-M 2-C 0.25-Q 1-output-debug-info**

Test options

- ☐ Use training set
- ☐ Supplied test set **Set...**
- ☒ Cross-validation **Folds** 10
- ☐ Percentage split **%** 66
- More options...**

Classifier output

PART decision list

Happiness Index = '(5.45-7.675)' AND
Global Slavery Index > 4.4: '(97.65-131.775)' (63.0/1.0)

Happiness Index = '(-inf-3.225)' AND
Human Flight = '(-inf-3.225)': '(-inf-63.525)' (20.0)

Happiness Index = '(7.675-inf)' AND
Legitimacy of the State = '(7.625-inf)' AND
Uneven Development = '(7.15-inf)': '(131.775-inf)' (29.0)

Global Warming Index = '(3.575-5.25)' AND
Demographic Pressures = '(3.9-5.9)': '(63.525-97.65)' (29.0/4.0)

Demographic Pressures = '(5.9-7.9)': '(97.65-131.775)' (12.0/1.0)

Happiness Index = '(3.225-5.45)' AND
Legitimacy of the State = '(3.075-5.35)': '(63.525-97.65)' (12.0/1.0)

Global Warming Index = '(-inf-3.575)': '(-inf-63.525)' (4.0)

Demographic Pressures = '(7.9-inf)': '(131.775-inf)' (6.0/2.0)

: '(63.525-97.65)' (3.0)

Number of Rules : 9

Result list (right-click for options)

- 14:16:03 - bayes.BayesNet
- 14:44:09 - rules.PART
- 14:47:26 - rules.PART

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize

Classifier

Choose
PART -M 2 -C 0.25 -Q 1 -output-debug-info

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Total

Start Stop

Result list (right-click for options)

14:16:03 - bayes.BayesNet
14:44:09 - rules.PART
14:47:26 - rules.PART

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      154           86.5169 %
Incorrectly Classified Instances    24           13.4831 %
Kappa statistic                    0.8038
Mean absolute error                 0.0691
Root mean squared error            0.2348
Relative absolute error            19.9253 %
Root relative squared error        56.4156 %
Total Number of Instances         178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC   ROC Area  PRC Area  Class
      -----  -
      0.875    0.013    0.913     0.875    0.894     0.878  0.977    0.954    '(-inf-63.525]'
      0.756    0.044    0.838     0.756    0.795     0.739  0.875    0.757    '(63.525-97.65]'
      0.900    0.122    0.857     0.900    0.878     0.775  0.918    0.855    '(97.65-131.775]'
      0.909    0.028    0.882     0.909    0.896     0.871  0.992    0.942    '(131.775-inf)'
Weighted Avg.  0.865    0.072    0.865     0.865    0.864     0.798  0.930    0.862

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
21  2  1  0  | a = '(-inf-63.525]'
 2 31  8  0  | b = '(63.525-97.65]'
 0  4 72  4  | c = '(97.65-131.775]'
 0  0  3 30  | d = '(131.775-inf)'

```

Year 2013

Discretize

We have Total as the decision feature. In the data set, the total column is sorted. Since we have to split the total into 4 intervals, we can split it as the levels Alert, Warning, Stable and Sustainable. We need to choose the split to get a maximum entropy gain for the decision feature considered.

Alert 35

Warning 78

Stable 42

Sustainable 23

Entropy Gain for Total

$$E(D) = -(((35/178) * \log_4(35/178)) + ((78/178) * \log_4(78/178)) + ((42/178) * \log_4(42/178)) + ((23/178) * \log_4(23/178)))$$

$$E(D) = -(0.196629213 * (-1.173225207) + 0.438202247 * (-0.595165606) + 0.235955056 * (-1.041708004) + 0.129213483 * (-1.476085737))$$

$$E(D) = 0.928019706$$

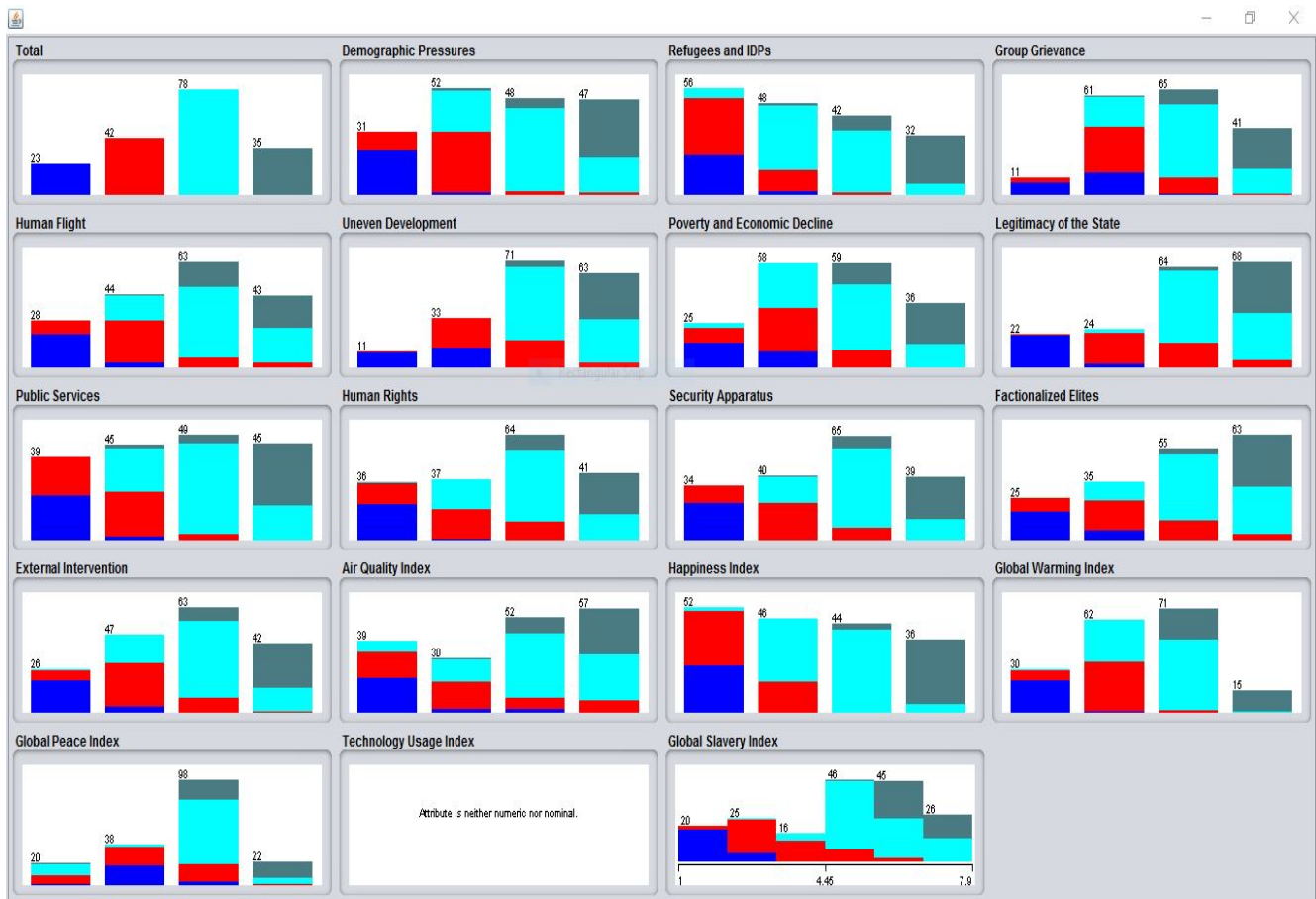
We have calculated the initial entropy and made the split based on the entropy gain of the split. We have chosen the split which has higher Entropy Gain and decided to have the total split into 4 intervals as below

(129.825 ... +infinity) (95.65 ... 129.825] (61.475 ... 95.65] (-infinity ... 61.475]

Classification using WEKA

Select Total as the decision feature and apply different algorithms to compare the results and choose the algorithm that gives the best number of correctly classified records.

Visualize all the features



Bayes – BayesNet Classifier

The screenshot shows the Weka Classifier window with the BayesNet classifier selected. The 'Test options' section on the left indicates 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane on the right displays the following results:

```

=== Summary ===
Correctly Classified Instances      163           91.573 %
Incorrectly Classified Instances    15            8.427 %
Kappa statistic                    0.8796
Mean absolute error                 0.041
Root mean squared error             0.1802
Relative absolute error             11.7326 %
Root relative squared error         43.1399 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
-----
0.957   0.019   0.880    0.957   0.917     0.905   0.997    0.988    '(-inf-61.475]'
0.833   0.015   0.946    0.833   0.886     0.857   0.992    0.972    '(61.475-95.65]'
0.923   0.050   0.935    0.923   0.929     0.874   0.992    0.990    '(95.65-129.825]'
0.971   0.035   0.872    0.971   0.919     0.900   0.997    0.987    '(129.825-inf)'
Weighted Avg.  0.916   0.035   0.918    0.916   0.915     0.879   0.993    0.985

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
22  1  0  0  | a = '(-inf-61.475]'
 3 35  4  0  | b = '(61.475-95.65]'
 0  1 72  5  | c = '(95.65-129.825]'
 0  0  1 34  | d = '(129.825-inf)'
  
```

Classifier Output

Correctly Classified Instances	163	91.573 %
Incorrectly Classified Instances	15	8.427 %
Kappa statistic	0.8796	
Mean absolute error	0.041	
Root mean squared error	0.1802	
Relative absolute error	11.7326 %	
Root relative squared error	43.1399 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.957	0.019	0.880	0.957	0.917	0.905	0.997	0.988	'(-inf-61.475]'
0.833	0.015	0.946	0.833	0.886	0.857	0.992	0.972	'(61.475-95.65]'
0.923	0.050	0.935	0.923	0.929	0.874	0.992	0.990	'(95.65-129.825]'
0.971	0.035	0.872	0.971	0.919	0.900	0.997	0.987	'(129.825-inf)'

Confusion Matrix

a	b	c	d	
22	1	0	0	a = '(-inf - 61.475]'
3	35	4	0	b = '(61.475 - 95.65]'
0	1	72	5	c = '(95.65 - 129.825]'
0	0	1	34	d = '(129.825 - inf)'

Bayes – Naive Bayes Classifier

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 16:01:13 - trees.RandomForest
- 16:03:56 - rules.PART
- 16:06:21 - bayes.BayesNet
- 16:07:09 - bayes.NaiveBayes

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      163           91.573 %
Incorrectly Classified Instances    15            8.427 %
Kappa statistic                    0.8798
Mean absolute error                 0.0441
Root mean squared error             0.1831
Relative absolute error             12.6409 %
Root relative squared error         43.8554 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
0.957   0.032   0.815   0.957   0.880   0.864   0.998   0.991   '(-inf-61.475]'
0.810   0.015   0.944   0.810   0.872   0.840   0.992   0.974   '(61.475-95.65]'
0.936   0.040   0.948   0.936   0.942   0.897   0.991   0.989   '(95.65-129.825]'
0.971   0.028   0.895   0.971   0.932   0.915   0.997   0.987   '(129.825-inf)'
Weighted Avg.   0.916   0.031   0.920   0.916   0.915   0.883   0.993   0.985

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
22  1  0  0  | a = '(-inf-61.475]'
 5 34  3  0  | b = '(61.475-95.65]'
 0  1 73  4  | c = '(95.65-129.825]'
 0  0  1 34  | d = '(129.825-inf)'

```

Classifier Output

Correctly Classified Instances 163 91.573 %

Incorrectly Classified Instances 15 8.427 %

Kappa statistic 0.8798

Mean absolute error 0.0441

Root mean squared error 0.1831

Relative absolute error 12.6409 %

Root relative squared error 43.8554 %

Total Number of Instances 178

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.957	0.032	0.815	0.957	0.880	0.864	0.998	0.991	'(-inf-61.475]'
0.810	0.015	0.944	0.810	0.872	0.840	0.992	0.974	'(61.475-95.65]'
0.936	0.040	0.948	0.936	0.942	0.897	0.991	0.989	'(95.65-129.825]'
0.971	0.028	0.895	0.971	0.932	0.915	0.997	0.987	'(129.825-inf)'

Confusion Matrix

a	b	c	d	
22	1	0	0	a = '(-inf - 61.475]'
5	34	3	0	b = '(61.475 - 95.65]'
0	1	73	4	c = '(95.65 - 129.825]'
0	0	1	34	d = '(129.825 - inf)'

Trees – Random Tree Classifier

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomTree-K 0-M 1.0-V 0.001-S 1

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 16:01:13 - trees.RandomForest
- 16:03:56 - rules.PART
- 16:06:21 - bayes.BayesNet
- 16:07:09 - bayes.NaiveBayes
- 16:07:51 - trees.HoeffdingTree
- 16:08:49 - trees.RandomTree

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      143      80.3371 %
Incorrectly Classified Instances    35      19.6629 %
Kappa statistic                    0.7163
Mean absolute error                 0.0959
Root mean squared error             0.2992
Relative absolute error             27.4697 %
Root relative squared error         71.6575 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.870    0.013    0.909     0.870    0.889     0.873    0.928     0.807    '(-inf-61.475]'
          0.714    0.066    0.769     0.714    0.741     0.665    0.839     0.656    '(61.475-95.65]'
          0.833    0.160    0.802     0.833    0.818     0.671    0.862     0.771    '(95.65-129.825]'
          0.800    0.056    0.778     0.800    0.789     0.736    0.877     0.714    '(129.825-inf)'
Weighted Avg.    0.803    0.098    0.804     0.803    0.803     0.709    0.868     0.738

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
20  3  0  0  a = '(-inf-61.475]'
 2 30  9  1  b = '(61.475-95.65]'
 0  6 65  7  c = '(95.65-129.825]'
 0  0  7 28  d = '(129.825-inf)'

```

Classifier Output

Correctly Classified Instances 143 80.3371 %

Incorrectly Classified Instances 35 19.6629 %

Kappa statistic 0.7163

Mean absolute error 0.0959

Root mean squared error 0.2992

Relative absolute error 27.4697 %

Root relative squared error 71.6575 %

Total Number of Instances 178

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.870	0.013	0.909	0.870	0.889	0.873	0.928	0.807	'(-inf-61.475]'
0.714	0.066	0.769	0.714	0.741	0.665	0.839	0.656	'(61.475-95.65]'
0.833	0.160	0.802	0.833	0.818	0.671	0.862	0.771	'(95.65-129.825]'
0.800	0.056	0.778	0.800	0.789	0.736	0.877	0.714	'(129.825-inf)'

Confusion Matrix

$$\begin{bmatrix}
 a & b & c & d \\
 20 & 3 & 0 & 0 \\
 2 & 30 & 9 & 1 \\
 0 & 6 & 65 & 7 \\
 0 & 0 & 7 & 28
 \end{bmatrix}
 \begin{aligned}
 a &= '(-inf - 61.475]' \\
 b &= '(61.475 - 95.65]' \\
 c &= '(95.65 - 129.825]' \\
 d &= '(129.825 - inf)'
 \end{aligned}$$

Trees – Random Forest Classifier

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -I 100 -num-slots 1 -K D-M 1.0 -V 0.001 -S 1

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Total

Start Stop

Result list (right-click for options)

16:01:13 - trees RandomForest

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      160           89.8876 %
Incorrectly Classified Instances    18           10.1124 %
Kappa statistic                    0.8536
Mean absolute error                 0.1034
Root mean squared error             0.1977
Relative absolute error             29.5977 %
Root relative squared error         47.3378 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.913   0.006   0.955     0.913   0.933     0.924   0.993    0.970    '(-inf-61.475]'
               0.810   0.029   0.895     0.810   0.850     0.808   0.980    0.883    '(61.475-95.65]'
               0.936   0.100   0.880     0.936   0.907     0.831   0.984    0.981    '(95.65-129.825]'
               0.914   0.021   0.914     0.914   0.914     0.893   0.995    0.980    '(129.825-inf)'
Weighted Avg.   0.899   0.056   0.900     0.899   0.898     0.850   0.986    0.956

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
21  2  0  0 | a = '(-inf-61.475]'
 1 34  7  0 | b = '(61.475-95.65]'
 0  2 73  3 | c = '(95.65-129.825]'
 0  0  3 32 | d = '(129.825-inf)'

```

Classifier Output

Correctly Classified Instances	160	89.8876 %
Incorrectly Classified Instances	18	10.1124 %
Kappa statistic	0.8536	
Mean absolute error	0.1034	
Root mean squared error	0.1977	
Relative absolute error	29.5977 %	
Root relative squared error	47.3378 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.913	0.006	0.955	0.913	0.933	0.924	0.993	0.970	'(-inf-61.475]'
0.810	0.029	0.895	0.810	0.850	0.808	0.980	0.883	'(61.475-95.65]'
0.936	0.100	0.880	0.936	0.907	0.831	0.984	0.981	'(95.65-129.825]'
0.914	0.021	0.914	0.914	0.914	0.893	0.995	0.980	'(129.825-inf)'

Confusion Matrix

a	b	c	d	
21	2	0	0	a = '(-inf - 61.475]'
1	34	7	0	b = '(61.475 - 95.65]'
0	2	73	3	c = '(95.65 - 129.825]'
0	0	3	32	d = '(129.825 - inf)'

Trees – Hoeffding Tree Classifier

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **HoeffdingTree** -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 16:01:13 - trees.RandomForest
- 16:03:56 - rules.PART
- 16:06:21 - bayes.BayesNet
- 16:07:09 - bayes.NaiveBayes
- 16:07:51 - trees.HoeffdingTree

Classifier output

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      164           92.1348 %
Incorrectly Classified Instances    14            7.8652 %
Kappa statistic                    0.8877
Mean absolute error                 0.0439
Root mean squared error             0.1826
Relative absolute error             12.5719 %
Root relative squared error         43.7226 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.957   0.026   0.846     0.957   0.898     0.884    0.998    0.991    '(-inf-61.475]'
          0.833   0.015   0.946     0.833   0.886     0.857    0.992    0.975    '(61.475-95.65]'
          0.936   0.040   0.948     0.936   0.942     0.897    0.992    0.990    '(95.65-129.825]'
          0.971   0.028   0.895     0.971   0.932     0.915    0.997    0.987    '(129.825-inf)'
Weighted Avg.    0.921   0.030   0.924     0.921   0.921     0.889    0.994    0.986

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
22  1  0  0  | a = '(-inf-61.475]'
 4 35  3  0  | b = '(61.475-95.65]'
 0 1 73  4  | c = '(95.65-129.825]'
 0 0 1 34  | d = '(129.825-inf)'
    
```

Classifier Output

Correctly Classified Instances	164	92.1348 %
Incorrectly Classified Instances	14	7.8652 %
Kappa statistic	0.8877	
Mean absolute error	0.0439	
Root mean squared error	0.1826	
Relative absolute error	12.5719 %	
Root relative squared error	43.7226 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.957	0.026	0.846	0.957	0.898	0.884	0.998	0.991	'(-inf-61.475]'
0.833	0.015	0.946	0.833	0.886	0.857	0.992	0.975	'(61.475-95.65]'
0.936	0.040	0.948	0.936	0.942	0.897	0.992	0.990	'(95.65-129.825]'
0.971	0.028	0.895	0.971	0.932	0.915	0.997	0.987	'(129.825-inf)'

Confusion Matrix

a	b	c	d	
22	1	0	0	a = '(-inf - 61.475]'
4	35	3	0	b = '(61.475 - 95.65]'
0	1	73	4	c = '(95.65 - 129.825]'
0	0	1	34	d = '(129.825 - inf)'

Comparison of Algorithms and Result

Since the percentage of correctly classified instances for Bayes Net, Naïve Bayes, Random Tree, Random Forest and Hoeffding Tree algorithms is 91.573, 91.573, 80.3371, 89.8876, 92.1348 respectively. This clearly shows that Hoeffding Tree algorithm is best used to classify as it has high accuracy when compared with other algorithms.

Determining Action Rules

We have determined the action rules by using Weka by selecting the Rules-PART filter as the classifier and run for the data with total as the decision attribute. The rules are logged in a word document and the result screenshot is shown below.

The screenshot displays the Weka software interface, specifically the 'Classifier' tab. The 'Choose' dropdown menu is set to 'PART-M 2-C 0.25-Q 1-output-debug-info'. The 'Test options' section on the left shows 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane on the right displays the 'PART decision list' with several rules. The 'Result list (right-click for options)' pane on the bottom left shows a list of results, including '16:03:56 - rules.PART'.

Classifier

Choose: PART-M 2-C 0.25-Q 1-output-debug-info

Test options

- ☐ Use training set
- ☐ Supplied test set (Set...)
- ☒ Cross-validation Folds: 10
- ☐ Percentage split %: 66
- More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 16:01:13 - trees.RandomForest
- 16:03:56 - rules.PART

Classifier output

PART decision list

Happiness Index = '(5.45-7.675)' AND
External Intervention = '(5.5-7.75)': '(95.65-129.825)' (26.0)

Happiness Index = '(7.675-inf)' AND
Global Slavery Index > 5.7: '(129.825-inf)' (34.0/2.0)

Global Warming Index = '(5.05-6.675)' AND
Global Slavery Index > 3.6 AND
Uneven Development = '(5.2-7.3)': '(95.65-129.825)' (22.0)

Demographic Pressures = '(-inf-3.7)' AND
Legitimacy of the State = '(-inf-2.825)': '(-inf-61.475)' (20.0)

Happiness Index = '(-inf-3.225)' AND
Human Flight = '(3.025-5.05)' AND
Technology Usage Index = '(3.35-5.5)': '(61.475-95.65)' (9.0)

Happiness Index = '(5.45-7.675)' AND
Group Grievance = '(7.75-inf)': '(95.65-129.825)' (6.0)

Happiness Index = '(-inf-3.225)' AND
Air Quality Index = '(7.65-inf)': '(61.475-95.65)' (6.0)

Happiness Index = '(5.45-7.675)' AND
Global Slavery Index <= 6.4: '(129.825-inf)' (3.0)

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize

Classifier

Choose
PART -M 2 -C 0.25 -Q 1 -output-debug-info

Test options

Classifier output

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds
☐ Percentage split %

(Nom) Total

Result list (right-click for options)

16:01:13 - trees.RandomForest
16:03:56 - rules.PART

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      146           82.0225 %
Incorrectly Classified Instances    32           17.9775 %
Kappa statistic                    0.7404
Mean absolute error                 0.0973
Root mean squared error             0.2736
Relative absolute error             27.8605 %
Root relative squared error         65.5277 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
      0.783    0.006    0.947     0.783    0.857     0.843    0.946     0.897     '(-inf-61.475]'
      0.667    0.066    0.757     0.667    0.709     0.628    0.905     0.713     '(61.475-95.65]'
      0.859    0.140    0.827     0.859    0.843     0.716    0.888     0.879     '(95.65-129.825]'
      0.943    0.056    0.805     0.943    0.868     0.837    0.952     0.749     '(129.825-inf)'
Weighted Avg.  0.820    0.089    0.822     0.820    0.818     0.736    0.912     0.816

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
18  5  0  0 | a = '(-inf-61.475]'
 1 28 12  1 | b = '(61.475-95.65]'
 0  4 67  7 | c = '(95.65-129.825]'
 0  0  2 33 | d = '(129.825-inf)'

```

Year 2014

Discretize

We have Total as the decision feature. In the data set, the total column is sorted. Since we have to split the total into 4 intervals, we can split it as the levels Alert, Warning, Stable and Sustainable. We need to choose the split to get a maximum entropy gain for the decision feature considered.

Alert 36

Warning 79

Stable 38

Sustainable 25

Entropy Gain for Total

$$E(D) = -(((36/178) * \log_4(36/178)) + ((79/178) * \log_4(79/178)) + ((38/178) * \log_4(38/178)) + ((25/178) * \log_4(25/178)))$$

$$E(D) = -(0.202247191 * (-1.152904215) + 0.443820225 * (-0.585976341) + 0.213483146 * (-1.113902959) + 0.140449438 * (-1.415938621))$$

$$E(D) = 0.929907082$$

We have calculated the initial entropy and made the split based on the entropy gain of the split. We have chosen the split which has higher Entropy Gain and decided to have the total split into 4 intervals as below

(129.3 ... +infinity)

(95.6 ... 129.3]

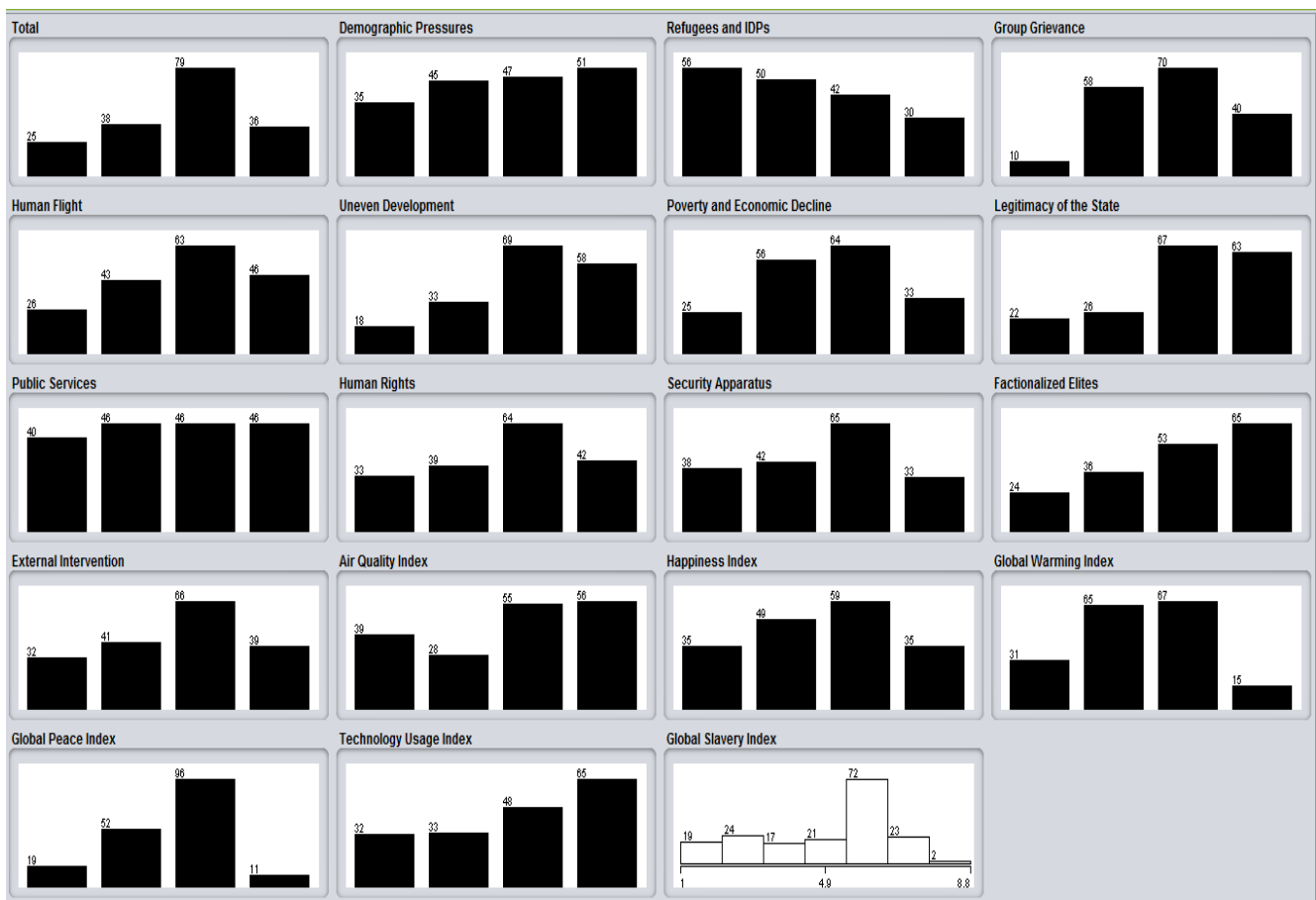
(61.9 ... 95.6]

(-infinity ... 61.9]

Classification using WEKA

Select Total as the decision feature and apply different algorithms to compare the results and choose the algorithm that gives the best number of correctly classified records.

Visualize all the features



Bayes - BayesNet Classifier

Classifier

Choose **BayesNet** - D - Q weka.classifiers.bayes.net.search.local.K2 -- P 1 - S BAYES - E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % **66**
More options...

Classifier output

```

Correctly Classified Instances      166           93.2584 %
Incorrectly Classified Instances    12            6.7416 %
Kappa statistic                    0.9034
Mean absolute error                 0.0391
Root mean squared error             0.1626
Relative absolute error             11.1841 %
Root relative squared error         38.9494 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
-----
0.960   0.020   0.889    0.960   0.923    0.911  0.997   0.981   '(-inf-61.9]'
0.895   0.014   0.944    0.895   0.919    0.898  0.989   0.968   '(61.9-95.6]'
0.949   0.040   0.949    0.949   0.949    0.909  0.991   0.987   '(95.6-129.3]'
0.917   0.021   0.917    0.917   0.917    0.896  0.996   0.984   '(129.3-inf)'
Weighted Avg.    0.933   0.028   0.933    0.933   0.933   0.904  0.992   0.982

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
24  1  0  0  a = '(-inf-61.9]'
 3 34  1  0  b = '(61.9-95.6]'
 0  1 75  3  c = '(95.6-129.3]'
 0  0  3 33  d = '(129.3-inf)'
    
```

Result list (right-click for options)

- 20:36:37 - trees.REPTree
- 20:39:38 - trees.REPTree
- 20:39:42 - trees.REPTree
- 20:39:47 - trees.REPTree
- 20:40:33 - trees.REPTree
- 20:40:39 - trees.REPTree
- 20:41:00 - trees.REPTree
- 20:42:19 - trees.REPTree
- 20:44:55 - trees.REPTree
- 20:46:48 - trees.HoeffdingTree
- 20:47:38 - bayes.BayesNet

Status

OK Log x0

Classifier Output

Correctly Classified Instances	166	93.2548 %
Incorrectly Classified Instances	12	6.7416 %
Kappa statistic	0.9034	
Mean absolute error	0.0391	
Root mean squared error	0.1626	
Relative absolute error	11.1841 %	
Root relative squared error	38.9494 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.960	0.020	0.889	0.960	0.923	0.911	0.997	0.981	'(-inf-61.9]'
0.895	0.014	0.944	0.895	0.919	0.898	0.989	0.968	'(61.9-95.6]'
0.949	0.040	0.949	0.949	0.949	0.909	0.991	0.987	'(95.6-129.3]'
0.917	0.021	0.917	0.917	0.917	0.896	0.996	0.984	'(129.3-inf)'

Confusion Matrix

a	b	c	d	
24	1	0	0	a = '(-inf - 61.9]'
3	34	1	0	b = '(61.9 - 95.6]'
0	1	75	3	c = '(95.6 - 129.3]'
0	0	3	33	d = '(129.3 - inf)'

Bayes - Naive Bayes Classifier

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Norm) Total

Result list (right-click for options)

- 20:39:38 - trees.REPtree
- 20:39:42 - trees.REPtree
- 20:39:47 - trees.REPtree
- 20:40:33 - trees.REPtree
- 20:40:39 - trees.REPtree
- 20:41:00 - trees.REPtree
- 20:42:19 - trees.REPtree
- 20:44:55 - trees.REPtree
- 20:46:48 - trees.HoeffdingTree
- 20:47:38 - bayes.BayesNet
- 20:48:28 - bayes.NaiveBayes

Classifier output

```


Summary
Correctly Classified Instances      164      92.1348 %
Incorrectly Classified Instances    14      7.8652 %
Kappa statistic                    0.8867
Mean absolute error                 0.043
Root mean squared error             0.1701
Relative absolute error             12.3219 %
Root relative squared error         40.73 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.960   0.026   0.857     0.960   0.906     0.891   0.997    0.982    '(-inf-61.9]'
0.816   0.007   0.969     0.816   0.886     0.863   0.989    0.970    '(61.9-95.6]'
0.962   0.061   0.927     0.962   0.944     0.898   0.991    0.989    '(95.6-129.3]'
0.917   0.021   0.917     0.917   0.917     0.896   0.996    0.985    '(129.3-inf)'
Weighted Avg.  0.921   0.036   0.924     0.921   0.921     0.889   0.993    0.983

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
24  1  0  0  a = '(-inf-61.9]'
 4 31  3  0  b = '(61.9-95.6]'
 0  0 76  3  c = '(95.6-129.3]'
 0  0  3 33  d = '(129.3-inf)'
    
```

Status

OK  x0

Classifier Output

Correctly Classified Instances	164	92.1348 %
Incorrectly Classified Instances	14	7.8652 %
Kappa statistic	0.8867	
Mean absolute error	0.043	
Root mean squared error	0.1701	
Relative absolute error	12.3219 %	
Root relative squared error	40.73 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.960	0.026	0.857	0.958	0.960	0.891	0.997	0.982	'(-inf-61.9]'
0.816	0.007	0.969	0.902	0.886	0.863	0.989	0.970	'(61.9-95.6]'
0.962	0.061	0.927	0.963	0.944	0.898	0.993	0.989	'(95.6-129.3]'
0.917	0.021	0.917	1.000	0.917	0.896	0.999	0.985	'(129.3-inf)'

Confusion Matrix

a	b	c	d	
24	1	0	0	a = '(-inf - 61.9]'
4	31	3	0	b = '(61.9 - 95.6]'
0	0	76	3	c = '(95.6 - 129.3]'
0	0	3	33	d = '(129.3 - inf)'

Trees – Random Tree Classifier

Classifier

Choose **RandomTree** -K 0-M 1.0-V 0.001-S 1

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66
 More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 20:00:00 - trees.RepTree
- 20:40:33 - trees.RepTree
- 20:40:39 - trees.RepTree
- 20:41:00 - trees.RepTree
- 20:42:19 - trees.RepTree
- 20:44:55 - trees.RepTree
- 20:46:48 - trees.HoeffdingTree
- 20:47:38 - bayes.BayesNet
- 20:48:28 - bayes.NaiveBayes
- 20:49:18 - trees.RandomForest
- 20:49:29 - trees.RandomForest
- 20:50:06 - trees.RandomTree

Classifier output

```

=== Summary ===
Correctly Classified Instances      145           81.4607 %
Incorrectly Classified Instances    33           18.5393 %
Kappa statistic                    0.7341
Mean absolute error                 0.0942
Root mean squared error             0.3018
Relative absolute error             26.9778 %
Root relative squared error         72.2796 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      -----
      0.920    0.020    0.885    0.920    0.902    0.886    0.949    0.812    '(-inf-61.9]'
      0.737    0.057    0.778    0.737    0.757    0.693    0.838    0.641    '(61.9-95.6]'
      0.823    0.141    0.823    0.823    0.823    0.681    0.850    0.763    '(95.6-129.3]'
      0.806    0.056    0.784    0.806    0.795    0.742    0.877    0.688    '(129.3-inf)'
Weighted Avg.    0.815    0.089    0.814    0.815    0.814    0.725    0.867    0.729

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
23  2  0  0  a = '(-inf-61.9]'
 3 28  7  0  b = '(61.9-95.6]'
 0  6 65  8  c = '(95.6-129.3]'
 0  0  7 29  d = '(129.3-inf)'
  
```

Status

OK Log x0

Classifier Output

Correctly Classified Instances	145	81.4607 %
Incorrectly Classified Instances	33	18.5393 %
Kappa statistic	0.7341	
Mean absolute error	0.0942	
Root mean squared error	0.3018	
Relative absolute error	26.9778 %	
Root relative squared error	72.2796 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.920	0.020	0.885	0.920	0.902	0.886	0.949	0.812	'(-inf-61.9]'
0.737	0.057	0.778	0.737	0.757	0.693	0.838	0.641	'(61.9-95.6]'
0.823	0.141	0.823	0.823	0.823	0.681	0.850	0.763	'(95.6-129.3]'
0.806	0.056	0.784	0.795	0.795	0.742	0.877	0.688	'(129.3-inf)'

Confusion Matrix

a	b	c	d	
23	2	0	0	a = '(-inf - 61.9]'
3	28	7	0	b = '(61.9 - 95.6]'
0	6	65	8	c = '(95.6 - 129.3]'
0	0	7	29	d = '(129.3 - inf)'

Trees – Random Forest Classifier

Classifier

Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds
☐ Percentage split %

(Nom) Total

Result list (right-click for options)

- 20:39:47 - trees.REPTree
- 20:40:33 - trees.REPTree
- 20:40:39 - trees.REPTree
- 20:41:00 - trees.REPTree
- 20:42:19 - trees.REPTree
- 20:44:55 - trees.REPTree
- 20:46:48 - trees.HoeffdingTree
- 20:47:38 - bayes.BayesNet
- 20:48:28 - bayes.NaiveBayes
- 20:49:18 - trees.RandomForest
- 20:49:29 - trees.RandomForest

Classifier output

```

=== Summary ===
Correctly Classified Instances      155           87.0787 %
Incorrectly Classified Instances    23           12.9213 %
Kappa statistic                    0.8116
Mean absolute error                0.1184
Root mean squared error            0.2141
Relative absolute error            33.9019 %
Root relative squared error        51.2675 %
Total Number of Instances         178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0.880    0.013    0.917    0.880    0.898    0.882    0.996    0.978    '(-inf-61.9]'
      0.789    0.029    0.882    0.789    0.833    0.793    0.975    0.932    '(61.9-95.6]'
      0.937    0.131    0.851    0.937    0.892    0.800    0.974    0.966    '(95.6-129.3]'
      0.806    0.028    0.879    0.806    0.841    0.804    0.988    0.960    '(129.3-inf)'
Weighted Avg.    0.871    0.072    0.872    0.871    0.870    0.811    0.980    0.959

=== Confusion Matrix ===
  a  b  c  d  <-- classified as
22  3  0  0  a = '(-inf-61.9]'
 2 30  6  0  b = '(61.9-95.6]'
 0  1 74  4  c = '(95.6-129.3]'
 0  0  7 29  d = '(129.3-inf)'

```

Status
OK x0

Classifier Output

Correctly Classified Instances	155	87.0787 %
Incorrectly Classified Instances	23	12.9213 %
Kappa statistic	0.8116	
Mean absolute error	0.1184	
Root mean squared error	0.2141	
Relative absolute error	33.9019 %	
Root relative squared error	51.2675 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.880	0.013	0.917	0.880	0.898	0.882	0.996	0.978	'(-inf-61.9]'
0.789	0.029	0.882	0.789	0.833	0.793	0.975	0.932	'(61.9-95.6]'
0.937	0.131	0.851	0.937	0.892	0.800	0.974	0.966	'(95.6-129.3]'
0.806	0.028	0.879	0.806	0.841	0.804	0.988	0.960	'(129.3-inf)'

Confusion Matrix

a	b	c	d	
22	3	0	0	a = '(-inf - 61.9]'
2	30	6	0	b = '(61.9 - 95.6]'
0	1	74	4	c = '(95.6 - 129.3]'
0	0	7	29	d = '(129.3 - inf)'

Trees – Hoeffding Tree Classifier

Classifier

Choose **HoeffdingTree** -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Norm) Total

Result list (right-click for options)

- 19:56:28 - trees.REPTree
- 20:36:37 - trees.REPTree
- 20:39:38 - trees.REPTree
- 20:39:42 - trees.REPTree
- 20:39:47 - trees.REPTree
- 20:40:33 - trees.REPTree
- 20:40:39 - trees.REPTree
- 20:41:00 - trees.REPTree
- 20:42:19 - trees.REPTree
- 20:44:55 - trees.REPTree
- 20:46:48 - trees.HoeffdingTree

Classifier output

```

=== Summary ===
Correctly Classified Instances      165          92.6966 %
Incorrectly Classified Instances    13           7.3034 %
Kappa statistic                    0.895
Mean absolute error                0.0427
Root mean squared error            0.1694
Relative absolute error            12.2344 %
Root relative squared error        40.5774 %
Total Number of Instances          178


=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
0.960   0.026   0.857     0.960   0.906     0.891   0.997    0.980    '(-inf-61.9]'
0.842   0.007   0.970     0.842   0.901     0.880   0.989    0.969    '(61.9-95.6]'
0.962   0.051   0.938     0.962   0.950     0.909   0.992    0.989    '(95.6-129.3]'
0.917   0.021   0.917     0.917   0.917     0.896   0.996    0.985    '(129.3-inf)'
Weighted Avg.  0.927   0.032   0.929     0.927   0.927     0.898   0.993    0.983

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
24  1  0  0  a = '(-inf-61.9]'
 4 32  2  0  b = '(61.9-95.6]'
 0  1 76  3  c = '(95.6-129.3]'
 0  0  3 33  d = '(129.3-inf)'
    
```

Status

OK  x0

Classifier Output

Correctly Classified Instances	165	95.5056 %
Incorrectly Classified Instances	13	4.4944 %
Kappa statistic	0.895	
Mean absolute error	0.0427	
Root mean squared error	0.1694	
Relative absolute error	12.2344 %	
Root relative squared error	40.5774 %	
Total Number of Instances	178	

TP	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.960	0.026	0.857	0.960	0.906	0.891	0.997	0.980	'(-inf-61.9]'
0.842	0.007	0.970	0.842	0.901	0.880	0.989	0.969	'(61.9-95.6]'
0.962	0.051	0.938	0.962	0.950	0.909	0.992	0.989	'(95.6-129.3]'
0.917	0.021	0.917	0.927	0.917	0.896	0.996	0.985	'(129.3-inf)'

Confusion Matrix

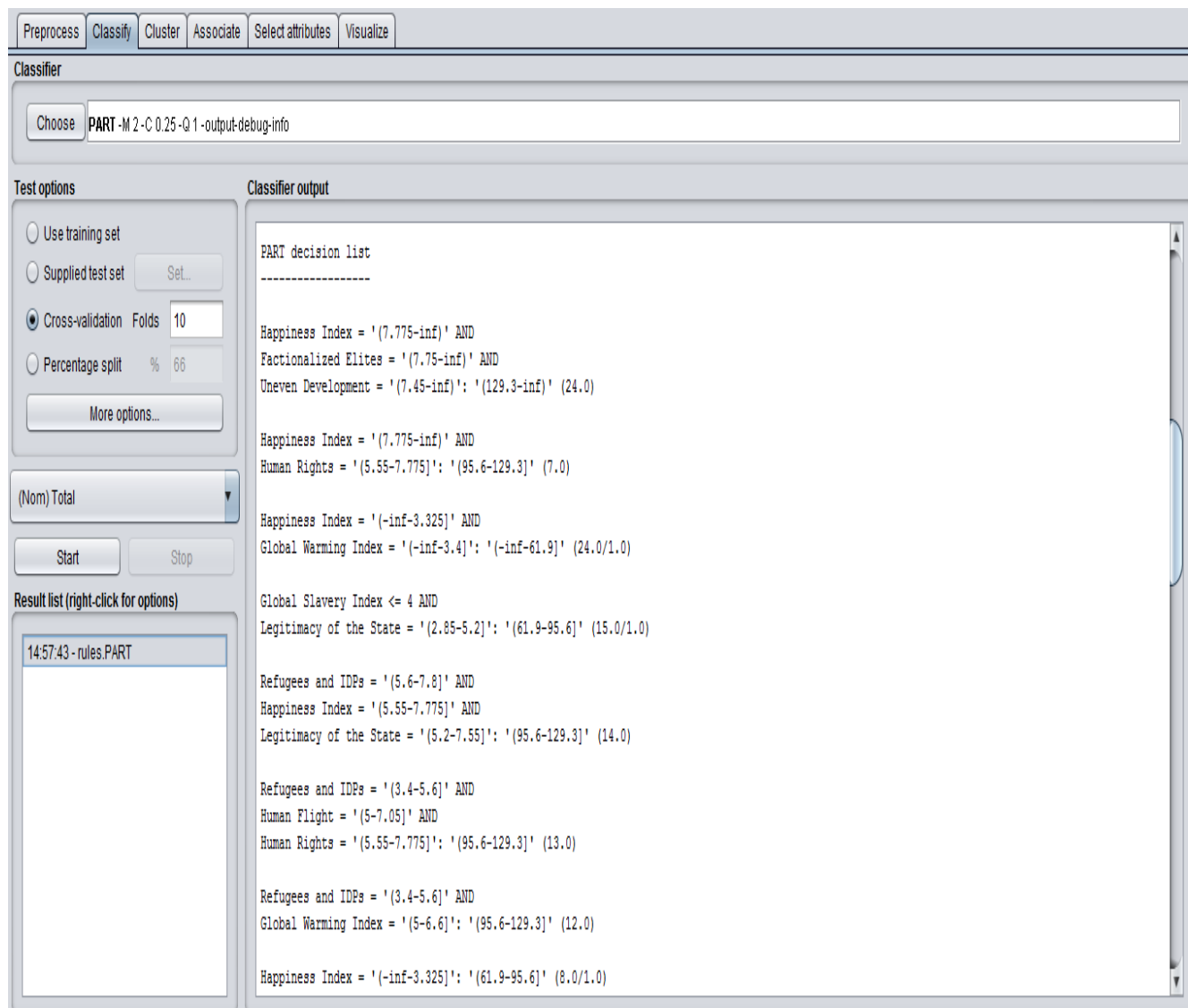
a	b	c	d	
24	1	0	0	a = '(-inf - 61.9]'
4	32	2	0	b = '(61.9 - 95.6]'
0	1	76	3	c = '(95.6 - 129.3]'
0	0	3	33	d = '(129.3 - inf)'

Comparison of Algorithms and Result

Since the percentage of correctly classified instances for Bayes Net, Naïve Bayes, Random Tree, Random Forest and Hoeffding Tree algorithms is 93.2548, 92.1348, 81.4607, 87.0787, 95.5056 respectively. This clearly shows that Hoeffding Tree algorithm is best used to classify as it has high accuracy when compared with other algorithms.

Determining Action Rules

We have determined the action rules by using Weka by selecting the Rules-PART filter as the classifier and run for the data with total as the decision attribute. The rules are logged in a word document and the result screenshot is shown below.



The screenshot displays the Weka software interface, specifically the 'Classifier' tab. The 'Choose' button is set to 'PART-M 2-C 0.25-Q 1-output-debug-info'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10. The 'Test on' dropdown is set to '(Nom) Total'. The 'Start' button is visible. The 'Classifier output' pane shows the 'PART decision list' with the following rules:

```
PART decision list
-----

Happiness Index = '(7.775-inf)' AND
Fractionalized Elites = '(7.75-inf)' AND
Uneven Development = '(7.45-inf)': '(129.3-inf)' (24.0)

Happiness Index = '(7.775-inf)' AND
Human Rights = '(5.55-7.775)': '(95.6-129.3)' (7.0)

Happiness Index = '(-inf-3.325)' AND
Global Warming Index = '(-inf-3.4)': '(-inf-61.9)' (24.0/1.0)

Global Slavery Index <= 4 AND
Legitimacy of the State = '(2.85-5.2)': '(61.9-95.6)' (15.0/1.0)

Refugees and IDPs = '(5.6-7.8)' AND
Happiness Index = '(5.55-7.775)' AND
Legitimacy of the State = '(5.2-7.55)': '(95.6-129.3)' (14.0)

Refugees and IDPs = '(3.4-5.6)' AND
Human Flight = '(5-7.05)' AND
Human Rights = '(5.55-7.775)': '(95.6-129.3)' (13.0)

Refugees and IDPs = '(3.4-5.6)' AND
Global Warming Index = '(5-6.6)': '(95.6-129.3)' (12.0)

Happiness Index = '(-inf-3.325)': '(61.9-95.6)' (8.0/1.0)
```

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

ChoosePART - M 2 - C 0.25 - Q 1 - output-debug-info

Test options

Use training set

Supplied test set

Cross-validation

Folds

10

Percentage split

%

66

More options...

(Nom) Total

Start

Stop

Result list (right-click for options)

14:57:43 - rules.PART

Classifier output

TIME TAKEN TO BUILD MODEL: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	138	77.5281 %
Incorrectly Classified Instances	40	22.4719 %
Kappa statistic	0.6779	
Mean absolute error	0.1268	
Root mean squared error	0.3135	
Relative absolute error	36.3205 %	
Root relative squared error	75.074 %	
Total Number of Instances	178	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.920	0.007	0.958	0.920	0.939	0.929	0.971	0.876	'(-inf-61.9]'
	0.711	0.107	0.643	0.711	0.675	0.582	0.835	0.512	'(61.9-95.6]'
	0.772	0.172	0.782	0.772	0.777	0.601	0.851	0.815	'(95.6-129.3]'
	0.750	0.049	0.794	0.750	0.771	0.716	0.941	0.843	'(129.3-inf)'
Weighted Avg.	0.775	0.110	0.780	0.775	0.777	0.666	0.883	0.765	

=== Confusion Matrix ===

a b c d <-- classified as

23 2 0 0 | a = '(-inf-61.9]'

1 27 9 1 | b = '(61.9-95.6]'

0 12 61 6 | c = '(95.6-129.3]'