ITCS 6162 - Knowledge Discovery in Databases

Analysis Report on Fragile States Index

Department of Computer Science

University of North Carolina Charlotte, NC, USA

Instructor

Dr. Zbigniew W. Ras

UNC CHARLOTTE
College of Computing and Informatics

Project Team

Group 14

Vishal Reddy Pachika (801259045)

Nithin Reddy Rupireddy (801257217)

Jyothi Ravipudi (801265491)

Aarsh Ghotra (801275191)

CONTENTS

1. PROJECT DESCRIPTION AND REQUIREMENTS

The main purpose of the project is to pre-process the Financial Stability Index (FSI) 2017 dataset and extract categorization and action rules. The Financial Stability Index (FSI) is a concept in which the United States presents a yearly report documenting the collapse. All sovereign states that are members of the United Nations and have sufficient data to determine their susceptibility to conflict or collapse are classified on this list.

We calculate how a country should shift from alert to a warning condition using Action-Rules. This focus also emphasizes the importance of each dataset attribute in generating the country's fragile state index.

FRAGILE STATE INDEX

Since 2005, the Fund for Peace and American Foreign Policy magazine have released an annual study called the Fragile States Index (FSI). The list ranks all sovereign governments that are members of the United Nations and have sufficient data to assess their vulnerability to conflict or collapse. Despite being recognized as sovereign by one or more other countries, Taiwan, the Palestinian Territories, Northern Cyprus, Kosovo, and Western Sahara are not named. The number of ratings for each of the 12 criteria determines the ranking. Each indicator is graded on a scale of 0 to 10, with 0 denoting the weakest (most stable) and 10 denoting the strongest (least stable), for a total of 0-120.

INDICATORS

Conflict risk indices are used to assess the current state of a country. The metrics provide a timely picture that can be compared to other snapshots in a time series to assess if things are becoming better or worse. The CAST system's and the Fragile States Index's metrics are listed below.

1. **Security Apparatus (C1):** It shows the size of a state's security threats. Attacks, bombings, the death rate after assaults, psychological oppression, and other examples are provided.

2. **Factionalized Elites (C2):** It depicts the state's divide along ethnic, social class, group, and race lines.

3. Group Dissatisfaction (C3): It highlights the social and political divisions that exist between various public meetings.

4. Economy (E1): It shows the depreciation of a country's currency. Unemployment rates, poverty levels, debt, business bankruptcies, and other factors can all play a part.

5. Economic Inequality (E2): This pointer displays an economic imbalance independent of a country's presentation.

6. Exodus of Humans and Brain Drain (E3): This represents people being uprooted (limited relocation when they are unable to progress economically) and the nation's sadness because of such development.

7. Legitimacy of the State (P1): It depicts the government's relationship with its citizens, as well as how accessible the government is to them.

8. Public Services (P2): It reflects the presence of many people-helping projects.

9. Human Rights (P3): It reflects how well the country's basic liberties and opportunities are maintained and preserved.

10. Tensions in the demographics (S1): It focuses on the disputes that the government has with the public, such as food security, access to common assets, and so on.

11. Refugees and Internally Displaced Persons (S2): It reveals the state's stress because of the state's limited ability to dislodge enormous social and political networks.

12. External Intervention (X1): It portrays the state's reactions to external events and the repercussions of those reactions.

2. EXTENDED FEATURES

This study proposes the discovery of new patterns and trends through the categorization of fragility within states. We expand on this study by introducing seven more factors that may aid in risk assessment and conflict detection. The following new features have been added:

1.Education (Mean Years of schooling (Years)): Average number of years of education received by people ages 25 or older, converted from education attainment levels using official duration of each level. By 2030, substantially increase the number of youths who have relevant skills, including technical and vocational skills, for employment, decent jobs, and entrepreneurship.

2.Unemployment: Percentage of the labor force population ages 15 and older that is not in paid employment or self-employed but is available for work and has taken steps to seek paid employment or self-employment. By 2030, achieve full and productive employment and decent work for all women and men, including for young people and persons with disabilities, and equal pay for work of equal value.

3.Health (Current health expenditure (% of GDP); Spending on healthcare goods and services, expressed as a percentage of GDP. It excludes capital health expenditures such as buildings, machinery, information technology and stocks of vaccines for emergency or outbreaks. Substantially increase health financing and the recruitment, development, training, and retention of the health workforce in developing countries, especially in least developed countries and small island developing states.

4.Population growth: The term “population growth” refers to an increase in the population of a nation. This information refers to the percentage increase from the previous year to the current year. Increased population growth contributes to increased inflation, military spending, unemployment, and other factors, all of which contribute to increased fragility.

5.Corruption: The Corruption Perceptions Index (CPI) is an index which ranks countries "by their perceived levels of public sector corruption, as determined by expert assessments and opinion surveys." The CPI generally defines corruption as an "abuse of entrusted power for private gain". The index is published annually by the non-governmental organization Transparency International since 1995.

6.Human Development Index (HDI): A composite index measuring average achievement in three basic dimensions of human development-a long and healthy life, knowledge, and a decent standard of living.
See Technical note 1 at the below URL
http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf for details on how the HDI is calculated.

3. DECISION ATTRIBUTES

According to the wiki page on the fragile index, the total score is calculated using 12 indicators and then used to identify countries.

The ranges defined in the Wikipedia page are:
- Alert (90.0 – 120.0)
- Warning (60.0 – 89.9)
- Stable (30.0-59.9)
- Sustainable (0-29.9)

After adding six additional qualities, the score was normalized as follows:
- The total was lowered from 180 to 120 (18 indications, each with a score ranging from 1 to 10).
- A score of 60 on 180, for example, would mean 60*120/180 = 40 on 120.

However, the six new indications introduced did not match to the existing twelve indicators, resulting in a substantially lower number than projected with the addition of six additional indicators. To account for the differences, the ranges were redefined; this range was derived by analyzing the original datasets and attempting to group nations in unstable conditions with identical numbers of countries.

The new ranges were

• Alert (>70.0)

• Warning (45.0 – 69.9)

• Stable (25.0-44.9)

• Sustainable (0-25)

This classification was applied to both the original and extended data sets.

4. DATA EXTRACTION

The values for these extended features have been extracted from various websites and added to the original FSI dataset. The data for year 2017 is compiled and stored as an Excel sheet for further processing & analysis.

Values for the extended features were gathered from various websites which are listed below:

- Education: https://hdr.undp.org/en/indicators/103006#
- Unemployment: https://hdr.undp.org/en/indicators/140606#
- Health: https://hdr.undp.org/en/indicators/137506
- Population: https://databank.worldbank.org/
- Corruption: https://www.transparency.org/en/cpi/2017
- Human Development Index: https://hdr.undp.org/en/indicators/137506

5. DATA PREPROCESSING

When data is pulled from several resources it is an inevitable fact that a huge amount unnecessary data will also be present, and it can cause an issue while performing desired tasks. For this reason, it is very important that we preprocess the data. Preprocessing is the method where we remove null values and few characters from the huge data to resolve upcoming problems. Once the data is preprocessed all the noisy data and the inconsistencies are removed before its execution. We can also check for missing values or data with the help of data preprocessing. So before using the data it is important that it is clean and organized as much as possible.

There are few characteristics needed for maintaining quality data:

1) Accuracy
2) Completeness

3) Timeliness
4) Consistency
5) Validated
6) Organized

All the above characteristics can be obtained by performing Data Preprocessing.

The data obtained from the different websites was not clean and hence some pre-processing was done before using it for classification. Following were the steps taken for Data Cleaning.

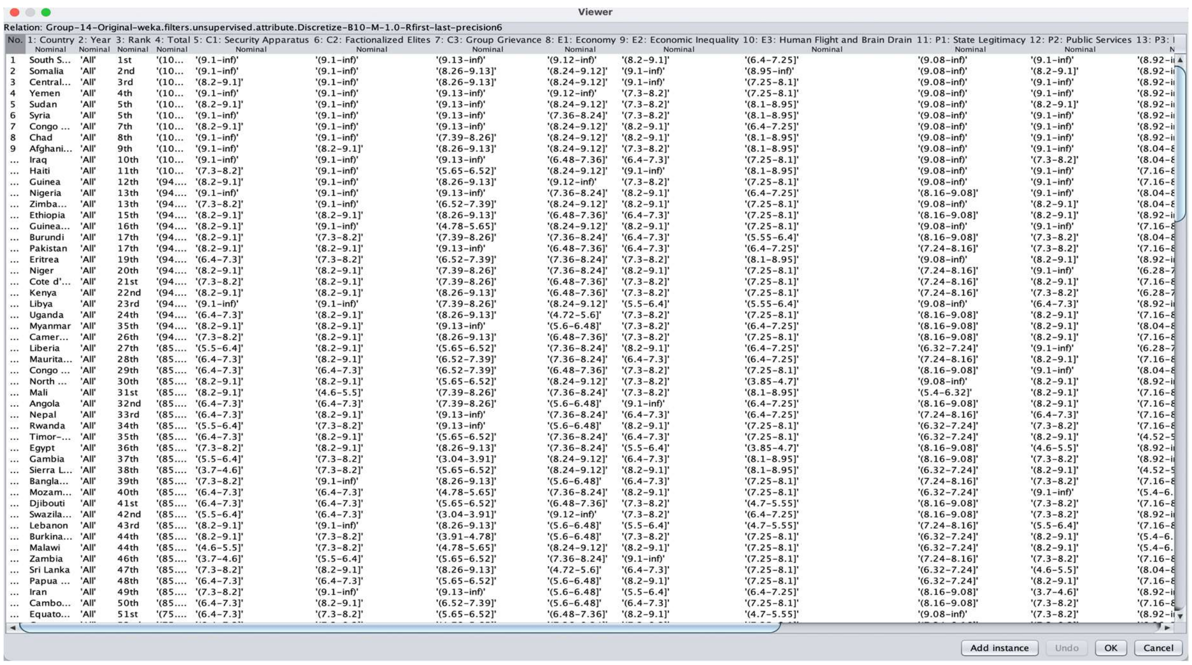


Figure 5.1: Representation of Preprocessed Data

- Special Characters like!, % etc. were removed so that the sheets can be parsed by WEKA tool.
- Missing numeric values were filled by their mean values.
- Rows having lot of empty values were removed from the analysis.
- Numeric values of Decision Variable Total were replaced by Nominal Values based on the following table values.
- Rounded the decimal values to 2 for better processing.
- Converted both original and extended excel files to CSV which then given as an input to WEKA software to filter the data.

6. DATA DISCRETIZATION USING WEKA

Discretization is a method where we try to reduce huge chunks of data into smaller values so that it is easy to manage the data.

Here we try to convert the real values into ordinal values or bins and the process is data discretization. We perform this when we make use of trees.

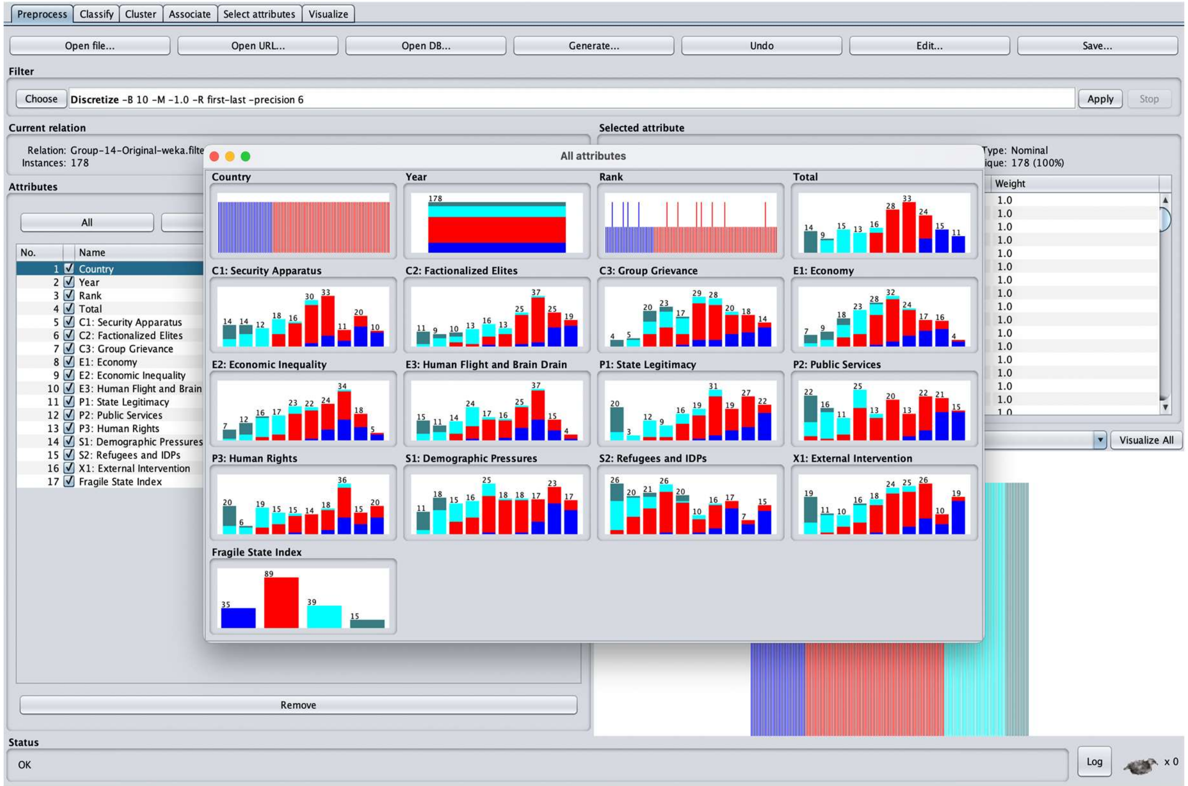


Figure 6.1: Visual Representation of Discretization

7. DATA CLASSIFICATION USING WEKA

Data Classification is a method of classifying data into discrete groupings. Categorization examines, interprets, and arranges material taken from massive databases, whether structured or unstructured; categorization analyzes, understands, and arranges information into numerous categories. It would be straightforward to decide which category a new item of data falls into and to add it to that category.

The following three classification algorithms were employed in this project.

1. Bayes Net Classifier:

A classifier which makes strong and independent assumptions on the data with the help of Bayes Theorem is Bayesian Network Classifier. Bayes Net Classifier is an independent feature model because it results in assuming that a presence/absence of a class feature is not relate with that of the presence/absence of any other feature. Bayes Networks are best for assuming that when an event occurred what is the likelihood that anyone of several possible causes was one of the contributing factors.

2. Random Forest Classifier:

Random Forest classifier is one if the simplest and easiest classifier to use. Since it is a supervised learning algorithm, it builds number of decision tress and merges them together and then it gets an accurate and stabilized prediction model.

There are two tasks performed while using this classifier:

1) Classification
2) Regression

In Classification tasks the output would be the class that has been selected by the highest number of trees, whereas, In Regression tasks the output would be the average prediction of the individual trees.

3. Randomizable Filter Classifier:

It is a class which is used for running an arbitrary classifier on data that is passed through an arbitrary filter. The classifier is based on the training data and test instances will be processed ahead using the filter without any changes or alteration in the structure. It is one of the simplest variants of filtered classifier that instantiates model with random projection. While we make use of Randomizable Filter as the base classifier each base classifier is built using a very different random number if seed. So, the final prediction is an average of predictions generated by individual base classifiers.

CLASSIFICATION USING ORIGINAL DATASET

We took 2017 data to perform the following experiment:

1. Bayes Net Classifier:

Input: Classify -> Choose -> BayesNet -> Apply

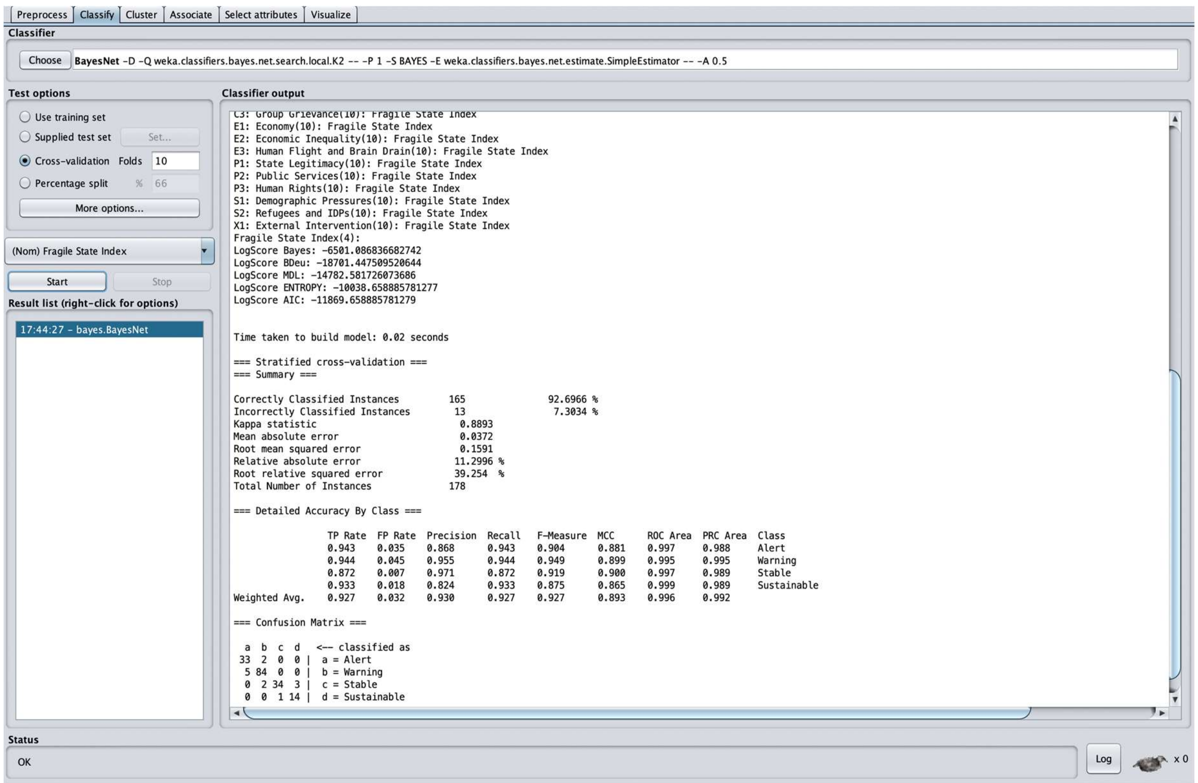


Figure 7.1: Applied Bayes Net Classifier on Original Data

Output:

=== Run information ===

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

Relation: Group-14-Original-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6

Instances: 178

Attributes: 17

Country

Year

Rank

Total

C1: Security Apparatus

C2: Factionalized Elites

C3: Group Grievance

E1: Economy

E2: Economic Inequality

E3: Human Flight and Brain Drain

P1: State Legitimacy

P2: Public Services

P3: Human Rights

S1: Demographic Pressures

S2: Refugees and IDPs

X1: External Intervention

Fragile State Index

Test mode: 10-fold cross-validation

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

Bayes Network Classifier

not using ADTree

#attributes=17 #classindex=16

Network structure (nodes followed by parents)

Country(178): Fragile State Index

Year(1): Fragile State Index

Rank(164): Fragile State Index

Total(10): Fragile State Index

C1: Security Apparatus(10): Fragile State Index

C2: Factionalized Elites(10): Fragile State Index

C3: Group Grievance(10): Fragile State Index

E1: Economy(10): Fragile State Index

E2: Economic Inequality(10): Fragile State Index

E3: Human Flight and Brain Drain(10): Fragile State Index

P1: State Legitimacy(10): Fragile State Index

P2: Public Services(10): Fragile State Index

P3: Human Rights(10): Fragile State Index

S1: Demographic Pressures(10): Fragile State Index

S2: Refugees and IDPs(10): Fragile State Index

X1: External Intervention(10): Fragile State Index

Fragile State Index(4):

LogScore Bayes: -6501.086836682742

LogScore BDeu: -18701.447509520644

LogScore MDL: -14782.581726073686

LogScore ENTROPY: -10038.658885781277

LogScore AIC: -11869.658885781279

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	165	92.6966 %
Incorrectly Classified Instances	13	7.3034 %
Kappa statistic	0.8893	
Mean absolute error	0.0372	
Root mean squared error	0.1591	
Relative absolute error	11.2996 %	
Root relative squared error	39.254 %	
Total Number of Instances	178	

=== Detailed Accuracy By Class ===

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

	0.943	0.035	0.868	0.943	0.904	0.881	0.997	0.988	Alert
	0.944	0.045	0.955	0.944	0.949	0.899	0.995	0.995	Warning
	0.872	0.007	0.971	0.872	0.919	0.900	0.997	0.989	Stable
	0.933	0.018	0.824	0.933	0.875	0.865	0.999	0.989	Sustainable
Weighted Avg.	0.927	0.032	0.930	0.927	0.927	0.893	0.996	0.992	

=== Confusion Matrix ===

```
 a  b  c  d   <-- classified as
33  2  0  0 |  a = Alert
 5 84  0  0 |  b = Warning
 0  2 34  3 |  c = Stable
 0  0  1 14 |  d = Sustainable
```

2. **Random Forest Classifier:**

Input: Classify -> Choose ->RandomForest -> Apply

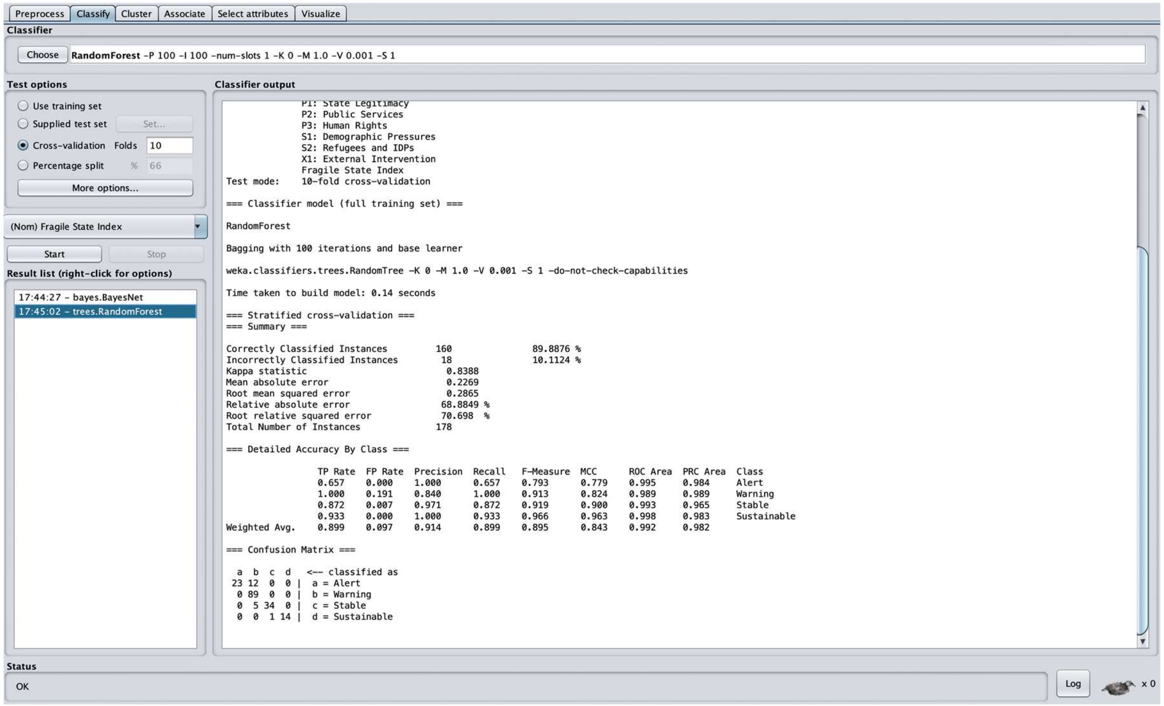


Figure 7.2: Applied Random Forest Classifier on Original Data

Output:

=== Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Relation: Group-14-Original-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6

Instances: 178

Attributes: 17

Country
Year
Rank
Total
C1: Security Apparatus
C2: Factionalized Elites
C3: Group Grievance
E1: Economy
E2: Economic Inequality
E3: Human Flight and Brain Drain
P1: State Legitimacy
P2: Public Services
P3: Human Rights
S1: Demographic Pressures
S2: Refugees and IDPs
X1: External Intervention
Fragile State Index

Test mode: 10-fold cross-validation

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.12 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	160	89.8876 %
Incorrectly Classified Instances	18	10.1124 %
Kappa statistic	0.8388	
Mean absolute error	0.2269	
Root mean squared error	0.2865	
Relative absolute error	68.8849 %	
Root relative squared error	70.698 %	
Total Number of Instances	178	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.657	0.000	1.000	0.657	0.793	0.779	0.995	0.984	Alert
	1.000	0.191	0.840	1.000	0.913	0.824	0.989	0.989	Warning
	0.872	0.007	0.971	0.872	0.919	0.900	0.993	0.965	Stable
	0.933	0.000	1.000	0.933	0.966	0.963	0.998	0.983	Sustainable
Weighted Avg.	0.899	0.097	0.914	0.899	0.895	0.843	0.992	0.982	

=== Confusion Matrix ===

```
  a  b  c  d   <-- classified as
 23 12  0  0 |  a = Alert
  0 89  0  0 |  b = Warning
```

0 5 34 0 | c = Stable

0 0 1 14 | d = Sustainable

3. Randomizable Filter Classifier:

Input: Classify -> Choose ->RandomizableFilter-> Apply

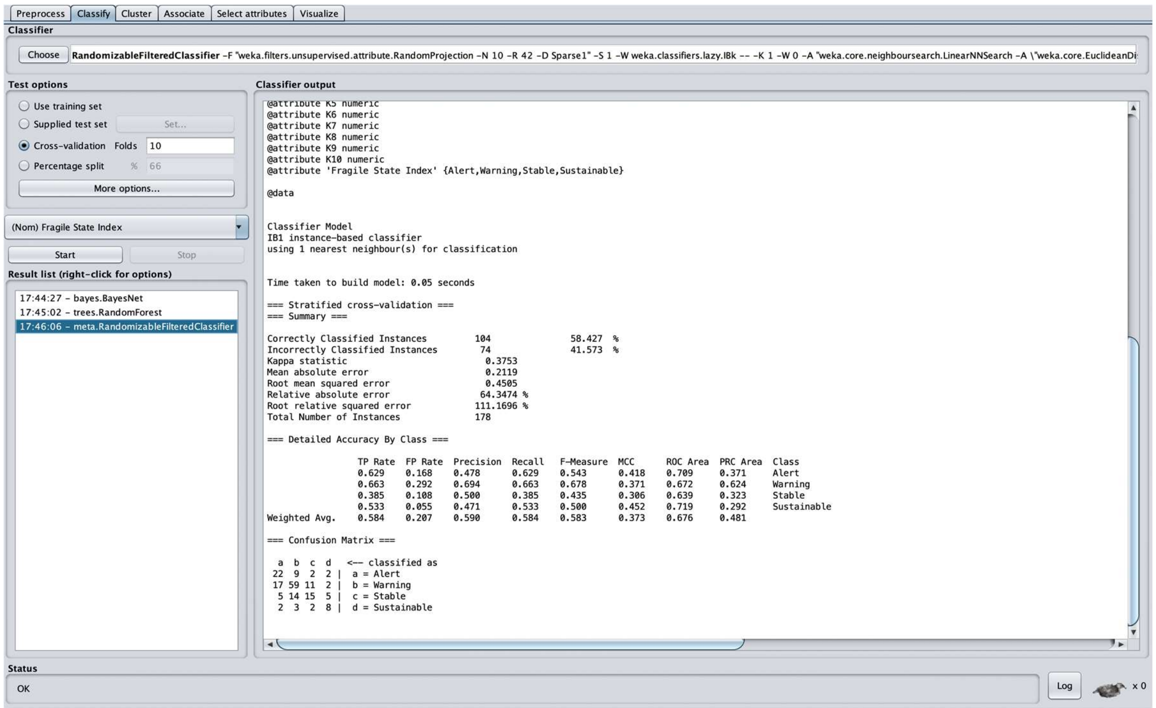


Figure 7.3: Applied Randomizable Filter Classifier on Original Data

Output:

=== Run information ===

Scheme: weka.classifiers.meta.RandomizableFilteredClassifier -F "weka.filters.unsupervised.attribute.RandomProjection -N 10 -R 42 -D Sparse1" -S 1 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: Group-14-Original-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6

Instances: 178

Attributes: 17

Country

Year

Rank

Total

C1: Security Apparatus

C2: Factionalized Elites

C3: Group Grievance

E1: Economy

E2: Economic Inequality

E3: Human Flight and Brain Drain

P1: State Legitimacy

P2: Public Services

P3: Human Rights

S1: Demographic Pressures

S2: Refugees and IDPs

X1: External Intervention

Fragile State Index

Test mode: 10-fold cross-validation

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

RandomizableFilteredClassifier using weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\"" on data filtered through weka.filters.unsupervised.attribute.RandomProjection -N 10 -R -990970977 -D Sparse1

Filtered Header

@relation Group-14-Original-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6-weka.filters.supervised.attribute.NominalToBinary-weka.filters.unsupervised.attribute.RandomProjection-N10-R-990970977-DSparse1

@attribute K1 numeric

@attribute K2 numeric

@attribute K3 numeric

@attribute K4 numeric

@attribute K5 numeric

@attribute K6 numeric

@attribute K7 numeric

@attribute K8 numeric

@attribute K9 numeric

@attribute K10 numeric

@attribute 'Fragile State Index' {Alert,Warning,Stable,Sustainable}

@data

Classifier Model

IB1 instance-based classifier

using 1 nearest neighbour(s) for classification

Time taken to build model: 0.06 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	104	58.427 %
Incorrectly Classified Instances	74	41.573 %
Kappa statistic	0.3753	
Mean absolute error	0.2119	
Root mean squared error	0.4505	
Relative absolute error	64.3474 %	
Root relative squared error	111.1696 %	
Total Number of Instances	178	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.629	0.168	0.478	0.629	0.543	0.418	0.709	0.371	Alert
	0.663	0.292	0.694	0.663	0.678	0.371	0.672	0.624	Warning
	0.385	0.108	0.500	0.385	0.435	0.306	0.639	0.323	Stable
	0.533	0.055	0.471	0.533	0.500	0.452	0.719	0.292	Sustainable
Weighted Avg.	0.584	0.207	0.590	0.584	0.583	0.373	0.676	0.481	

=== Confusion Matrix ===

```
 a  b  c  d   <-- classified as
22  9  2  2 |  a = Alert
17 59 11  2 |  b = Warning
 5 14 15  5 |  c = Stable
 2  3  2  8 |  d = Sustainable
```

CLASSIFICATION USING EXTENDED DATASET

We took 2017 data and added six new features to perform the following experiment:

1. **Bayes Net Classifier:**

   Input: Classify -> Choose -> BayesNet -> Apply

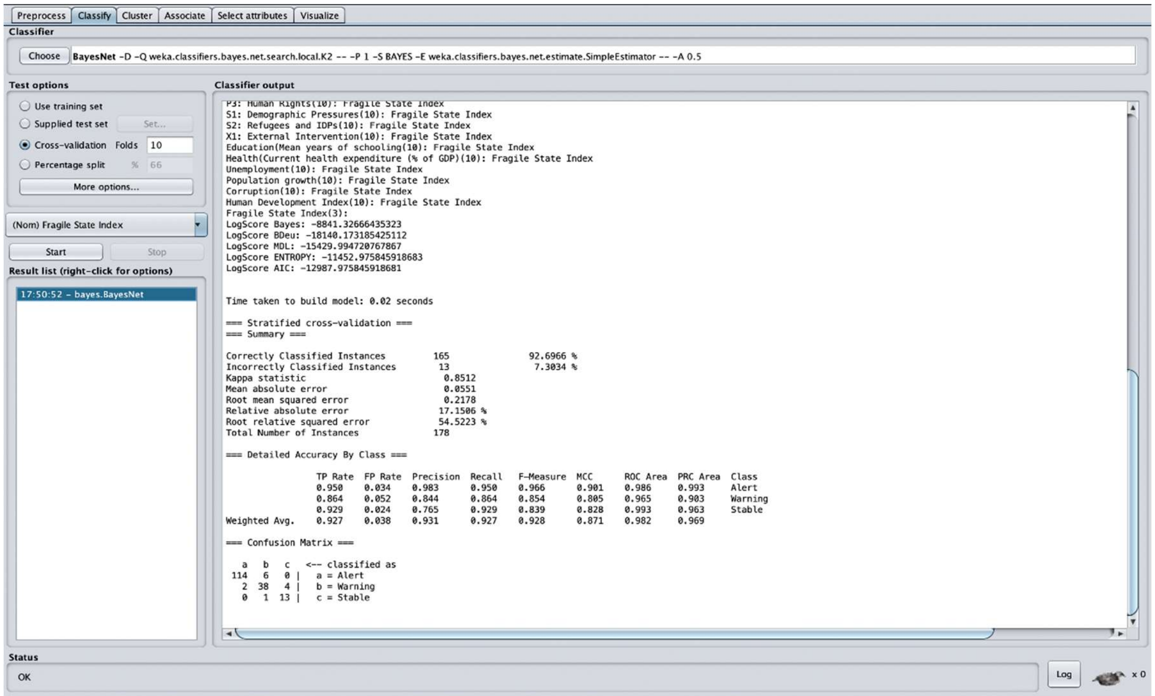


Figure 7.4: Applied Bayes Net Classifier on Extended Data

Output:

=== Run information ===

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

Relation: Group-14-Extended-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6

Instances: 178

Attributes: 23

Country

Year

Rank

Total

C1: Security Apparatus

C2: Factionalized Elites

C3: Group Grievance

E1: Economy

E2: Economic Inequality

E3: Human Flight and Brain Drain

P1: State Legitimacy

P2: Public Services

P3: Human Rights

S1: Demographic Pressures

S2: Refugees and IDPs

X1: External Intervention

Education(Mean years of schooling

Health(Current health expenditure (% of GDP)

Unemployment

Population growth

Corruption

Human Development Index

Fragile State Index

Test mode: 10-fold cross-validation

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

Bayes Network Classifier

not using ADTree

#attributes=23 #classindex=22

Network structure (nodes followed by parents)

Country(178): Fragile State Index

Year(1): Fragile State Index

Rank(164): Fragile State Index

Total(10): Fragile State Index

C1: Security Apparatus(10): Fragile State Index

C2: Factionalized Elites(10): Fragile State Index

C3: Group Grievance(10): Fragile State Index

E1: Economy(10): Fragile State Index

E2: Economic Inequality(10): Fragile State Index

E3: Human Flight and Brain Drain(10): Fragile State Index

P1: State Legitimacy(10): Fragile State Index

P2: Public Services(10): Fragile State Index

P3: Human Rights(10): Fragile State Index

S1: Demographic Pressures(10): Fragile State Index

S2: Refugees and IDPs(10): Fragile State Index

X1: External Intervention(10): Fragile State Index

Education(Mean years of schooling(10): Fragile State Index

Health(Current health expenditure (% of GDP)(10): Fragile State Index

Unemployment(10): Fragile State Index

Population growth(10): Fragile State Index

Corruption(10): Fragile State Index

Human Development Index(10): Fragile State Index

Fragile State Index(3):

LogScore Bayes: -8841.32666435323

LogScore BDeu: -18140.173185425112

LogScore MDL: -15429.994720767867

LogScore ENTROPY: -11452.975845918683

LogScore AIC: -12987.975845918681

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	165	92.6966 %
Incorrectly Classified Instances	13	7.3034 %
Kappa statistic	0.8512	
Mean absolute error	0.0551	
Root mean squared error	0.2178	
Relative absolute error	17.1506 %	
Root relative squared error	54.5223 %	

Total Number of Instances 178

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.950	0.034	0.983	0.950	0.966	0.901	0.986	0.993	Alert
	0.864	0.052	0.844	0.864	0.854	0.805	0.965	0.903	Warning
	0.929	0.024	0.765	0.929	0.839	0.828	0.993	0.963	Stable
Weighted Avg.	0.927	0.038	0.931	0.927	0.928	0.871	0.982	0.969	

=== Confusion Matrix ===

a b c <-- classified as

114 6 0 | a = Alert

2 38 4 | b = Warning

0 1 13 | c = Stable

2. **Random Forest Classifier:**

Input: Classify -> Choose ->RandomForest -> Apply

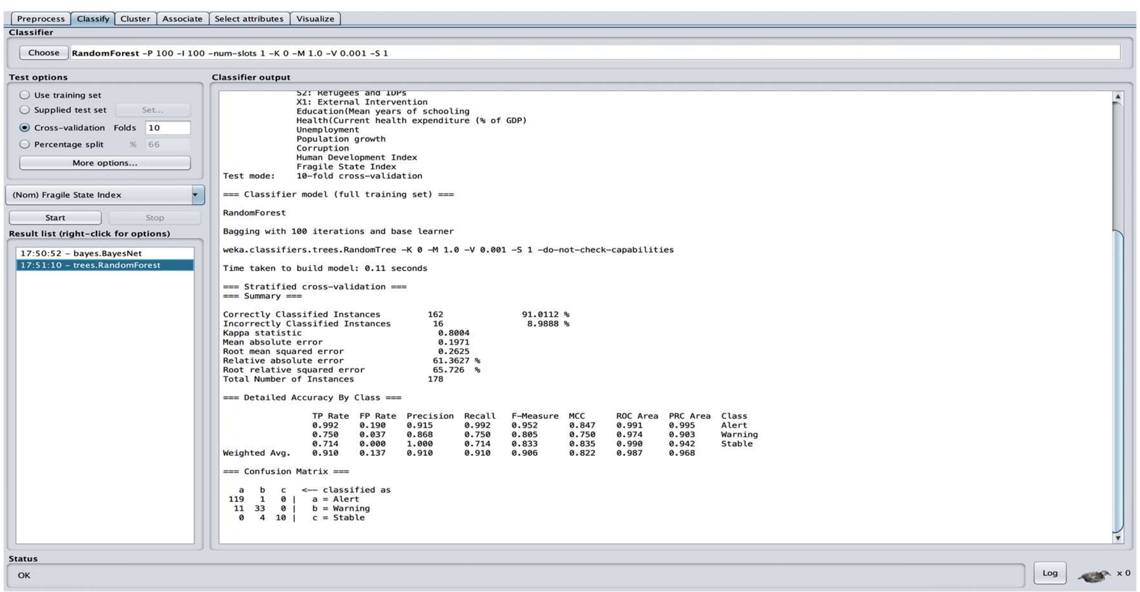


Figure 7.5: Applied Random Forest Classifier on Extended Data

Output:

=== Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Relation: Group-14-Extended-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6

Instances: 178

Attributes: 23

Country

Year

Rank

Total

C1: Security Apparatus

C2: Factionalized Elites

C3: Group Grievance

E1: Economy

E2: Economic Inequality

E3: Human Flight and Brain Drain

P1: State Legitimacy

P2: Public Services

P3: Human Rights

S1: Demographic Pressures

S2: Refugees and IDPs

X1: External Intervention

Education(Mean years of schooling

Health(Current health expenditure (% of GDP)

Unemployment

Population growth

Corruption

Human Development Index

Fragile State Index

Test mode: 10-fold cross-validation

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	162	91.0112 %
Incorrectly Classified Instances	16	8.9888 %
Kappa statistic	0.8004	
Mean absolute error	0.1971	
Root mean squared error	0.2625	
Relative absolute error	61.3627 %	
Root relative squared error	65.726 %	
Total Number of Instances	178	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.992	0.190	0.915	0.992	0.952	0.847	0.991	0.995	Alert
	0.750	0.037	0.868	0.750	0.805	0.750	0.974	0.903	Warning
	0.714	0.000	1.000	0.714	0.833	0.835	0.990	0.942	Stable
Weighted Avg.	0.910	0.137	0.910	0.910	0.906	0.822	0.987	0.968	

=== Confusion Matrix ===

```
  a   b   c   <-- classified as
119   1   0 |   a = Alert
 11  33   0 |   b = Warning
  0   4  10 |   c = Stable
```

3. **Randomizable Filter Classifier:**

Input: Classify -> Choose ->RandomizableFilter-> Apply

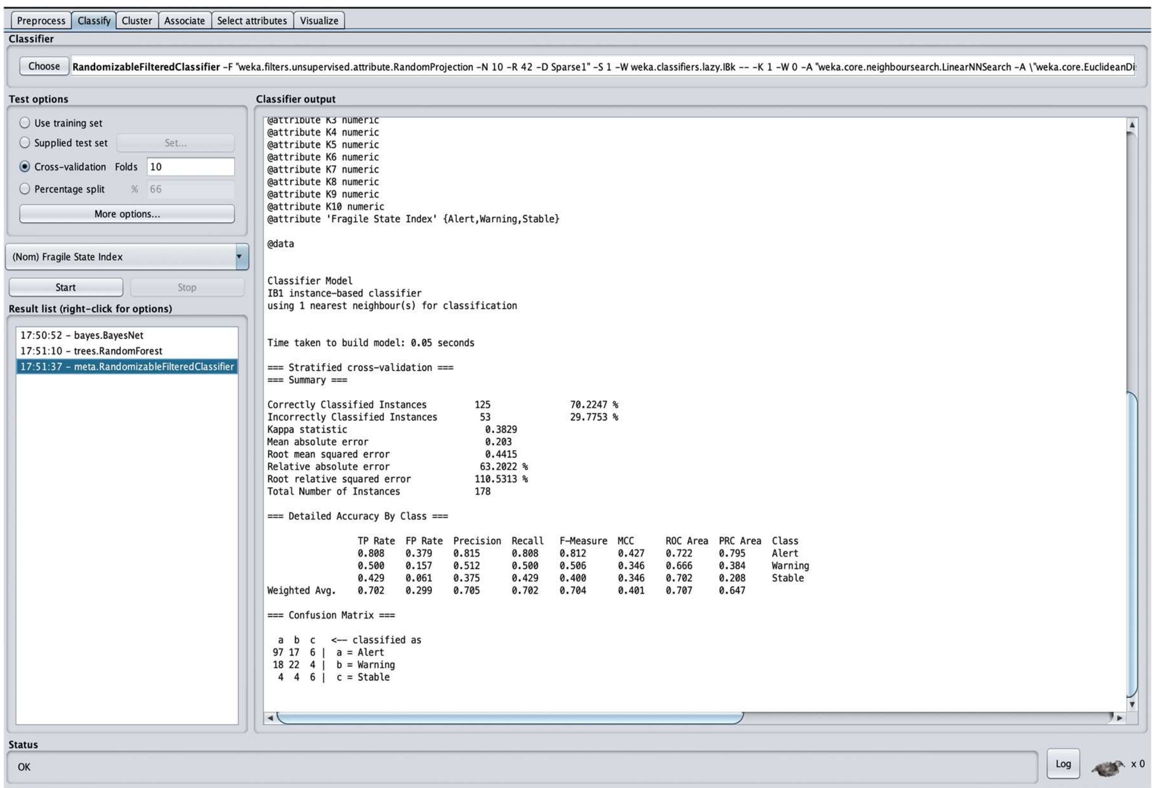


Figure 7.6: Applied Randomizable Filter Classifier on Extended Data

Output:

=== Run information ===

Scheme: weka.classifiers.meta.RandomizableFilteredClassifier -F "weka.filters.unsupervised.attribute.RandomProjection -N 10 -R 42 -D Sparse1" -S 1 -W weka.classifiers.lazy.IBk -- -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: Group-14-Extended-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6

Instances: 178

Attributes: 23

Country

Year

Rank

Total

C1: Security Apparatus

C2: Factionalized Elites

C3: Group Grievance

E1: Economy

E2: Economic Inequality

E3: Human Flight and Brain Drain

P1: State Legitimacy

P2: Public Services

P3: Human Rights

S1: Demographic Pressures

S2: Refugees and IDPs

X1: External Intervention

Education(Mean years of schooling

Health(Current health expenditure (% of GDP)

Unemployment

Population growth

Corruption

Human Development Index

Fragile State Index

Test mode: 10-fold cross-validation

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

RandomizableFilteredClassifier using weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\"" on data filtered through weka.filters.unsupervised.attribute.RandomProjection -N 10 -R -528505554 -D Sparse1

Filtered Header
@relation Group-14-Extended-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6-weka.filters.supervised.attribute.NominalToBinary-weka.filters.unsupervised.attribute.RandomProjection-N10-R-528505554-DSparse1

@attribute K1 numeric
@attribute K2 numeric
@attribute K3 numeric
@attribute K4 numeric
@attribute K5 numeric
@attribute K6 numeric
@attribute K7 numeric
@attribute K8 numeric
@attribute K9 numeric
@attribute K10 numeric
@attribute 'Fragile State Index' {Alert,Warning,Stable}

@data

Classifier Model
IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	125	70.2247 %
Incorrectly Classified Instances	53	29.7753 %
Kappa statistic	0.3829	
Mean absolute error	0.203	
Root mean squared error	0.4415	
Relative absolute error	63.2022 %	
Root relative squared error	110.5313 %	
Total Number of Instances	178	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.808	0.379	0.815	0.808	0.812	0.427	0.722	0.795	Alert
	0.500	0.157	0.512	0.500	0.506	0.346	0.666	0.384	Warning
	0.429	0.061	0.375	0.429	0.400	0.346	0.702	0.208	Stable
Weighted Avg.	0.702	0.299	0.705	0.702	0.704	0.401	0.707	0.647	

=== Confusion Matrix ===

```
  a  b  c   <-- classified as
 97 17  6 |  a = Alert
 18 22  4 |  b = Warning
  4  4  6 |  c = Stable
```

8. GENERATION OF ACTION RULES USING LISP MINER

ACTION RULES:

Action rules are applied to a database in which the characteristics are classified as either flexible or stable. To reclassify a set of objects into a new decision class, flexible characteristics are required. We utilize the Lisp Miner program to construct action rules for the Fragile State Index 2017 data set.

LISP MINER:

The LISP-Miner system is an academic data mining software program created at the University of Economics in Prague1. It is a project that focuses on mining various types of association rules from categorical data. LISP-Miner uses multiple data mining processes to create a broader variety of different sorts of links between the left and right sides of a rule. For this project, we are extracting action rules utilizing the Ac4ft-Miner data mining technique. Ac4ft-Miner identifies rules that outline which actions should be made to enhance the stated state. It does so by analyzing the relationships between the data supplied as input.

Lisp Miner's action rule discovery methodology is as follows:

LISP-Miner mines for many sorts of knowledge patterns using a variety of GUHA processes. This system consists of ten distinct data mining procedures, four of which are based on the original GUHA method ASSOC and the others were created during the system's development.

The 4ft-Miner process seeks knowledge patterns, which may be seen as 4ft association rules of the kind

$$\phi \approx \psi/\gamma$$

where ϕ (antecedent), ψ (succedent) and γ (condition) are cedents and $\approx$ is a quantifier that is applied to the subset of samples that fulfill the condition.

Dealing with LISP-Miner is more difficult than working with other data mining systems since it is delivered as a set of executables that must be called by the user. Each mining technique in the LISP-Miner program makes use of a variety of processing modules, including:

1. **LMAdmin:** The LMAdmin module is the first to be utilized. The main purpose of this module is to connect the meta-data base to the examined data. The idea of meta-data allows for the storing

and reuse of task inputs as well as analysis outcomes. A database is used to hold both data and meta-data. This is a mandatory step that must be completed prior to conducting any analysis.

2. **LMDataSource:** This module includes several data transformation and preparation techniques that may be used to choose characteristics for a specific data mining operation, produce derived attributes, or discretize numeric attributes.
3. **Data processing (Task):** This is the task module that analyzes the data and produces the task by utilizing the relevant xxxTask module.
4. **Data interpretation (Result):** The Result module is responsible for displaying and evaluating the results. When the relevant xxxResult module is executed, this is used to show, sort, or choose the rules that are created.

PROPERTIES:

We have classified our attributes into three categories:

• Stable Attributes

• Flexible Attributes

• Decision Attributes

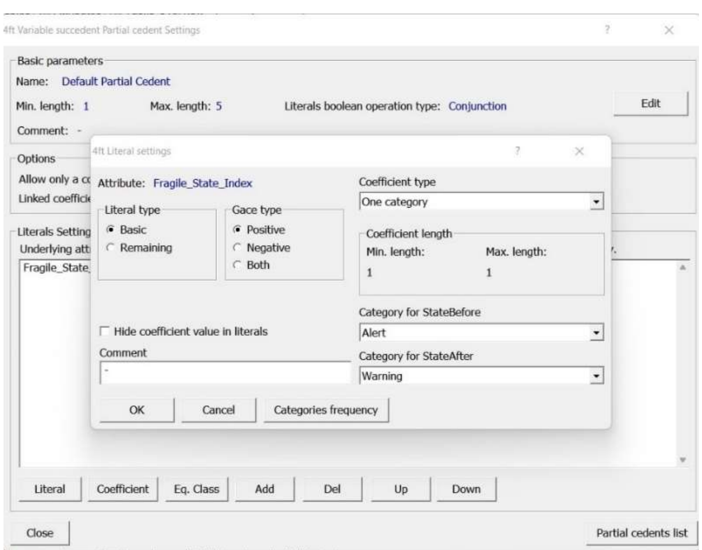
We choose the following stable properties:

• **Antecedent stable part:** We distribute all the consistent ascribes to this community.

• **Antecedent variable part:** This collection contains all the adaptable ascribes.

• **Succedent variable part:** For this set, we assigned the following decision variable.

- o Attribute type = nominal
- o Coefficient type = one category



Quantifiers :

• a(BASE)Before : 2

• a(BASE)After : 2

Figure 8.1: Changing the Coefficient and Category for Succedent Variable

9. LISP MINER SCREENSHOTS

Step 1: The initial stage is to get a dataset, which is subsequently modified by adding six new attributes.

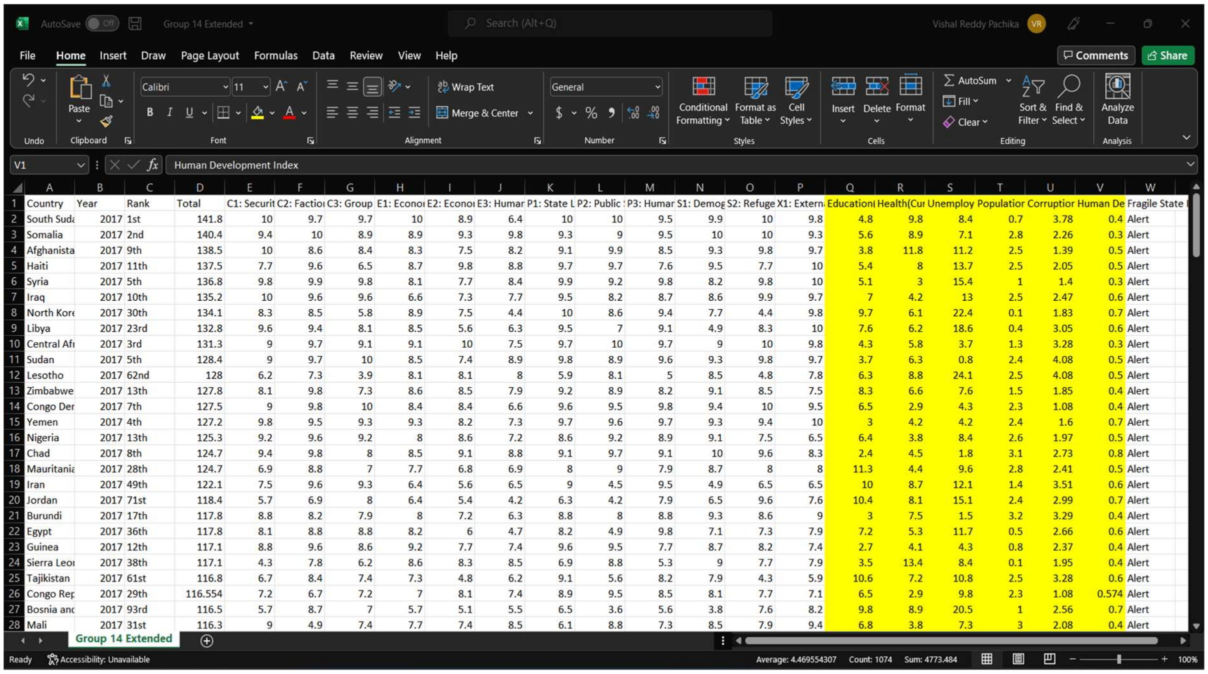


Figure 9.1: Extended Dataset after adding additional six features

Step 2: Import the dataset into the LISP miner program by clicking the 'New from TXT' button. When importing data via text file, the file must be of the text or csv format that processes all 178 rows, one for each country.

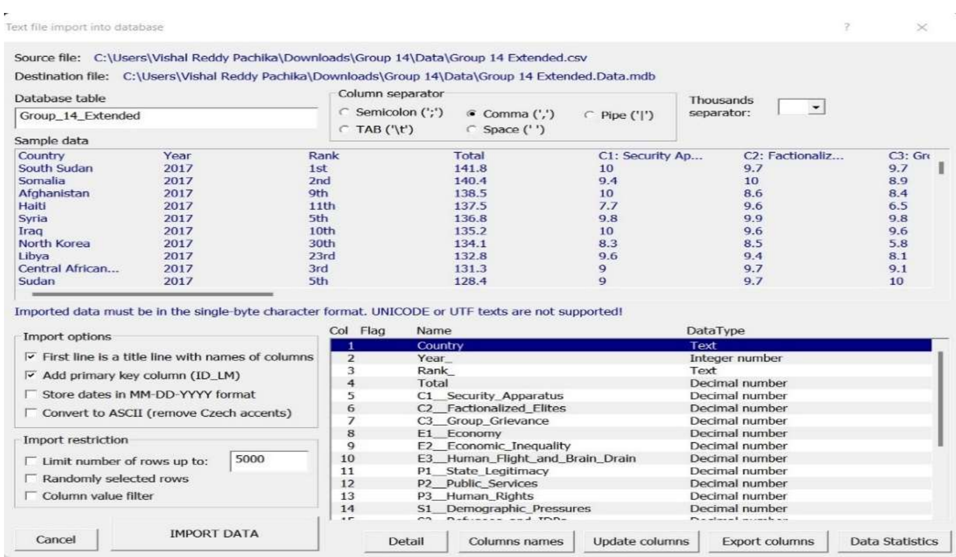


Figure 9.2: Extended data of 2017 imported in LISP Miner to perform further analysis.

Step 3: After the file is loaded, it can be viewed in the lisp miner software in the table as the dataset is converted to a database in the previous step.

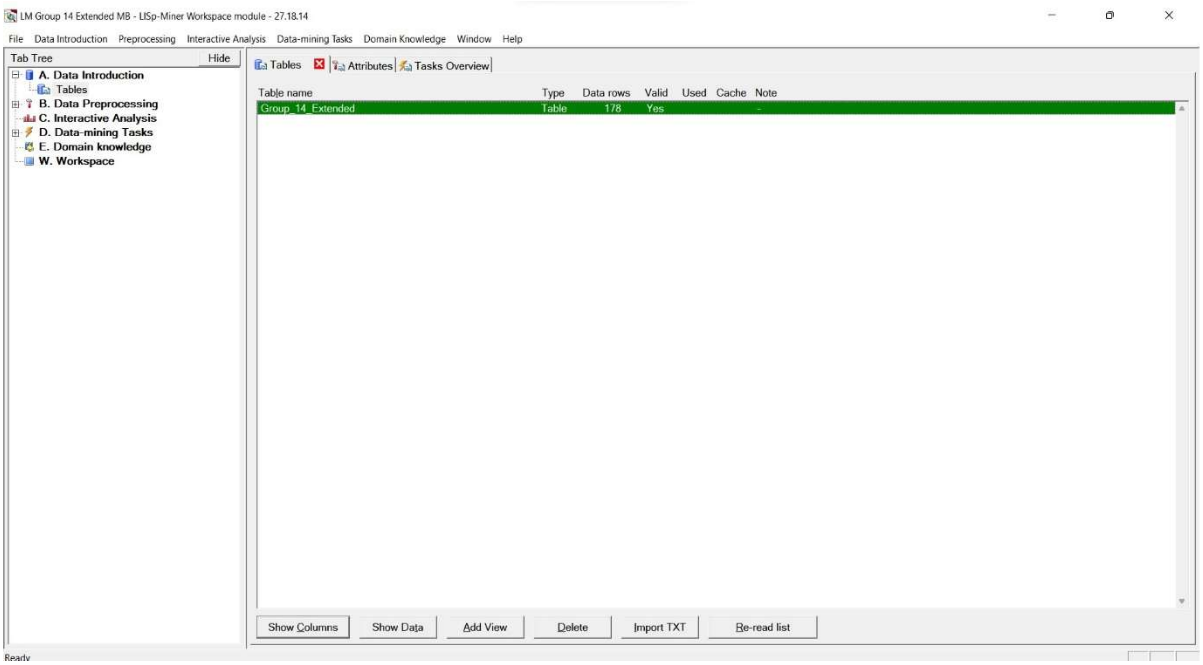


Figure 9.3: Dataset is converted to a database .mdb file.

Step 4: The entire database attributes can be seen in the columns sub section of Data Introduction in the Lisp Miner tool. Now, select all the columns except ID and click on create attribute.

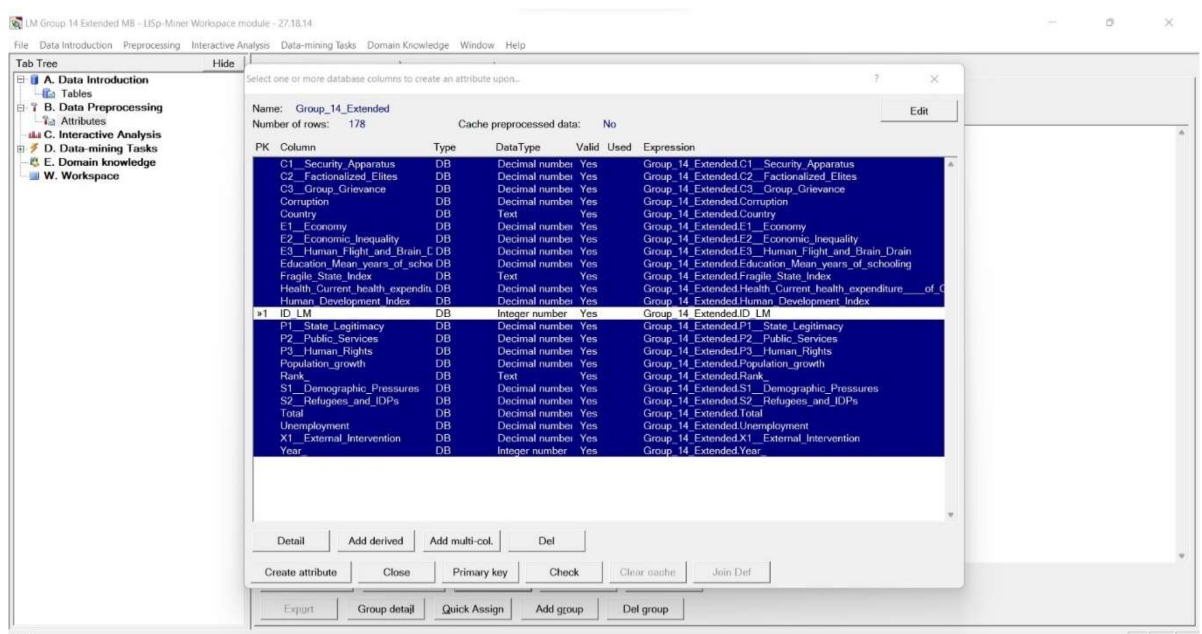


Figure 9.4: Selecting Attributes except system generated ID.

Now, Change 'f'-bound for interval boundaries in Float values from 0.01 to 0.001.

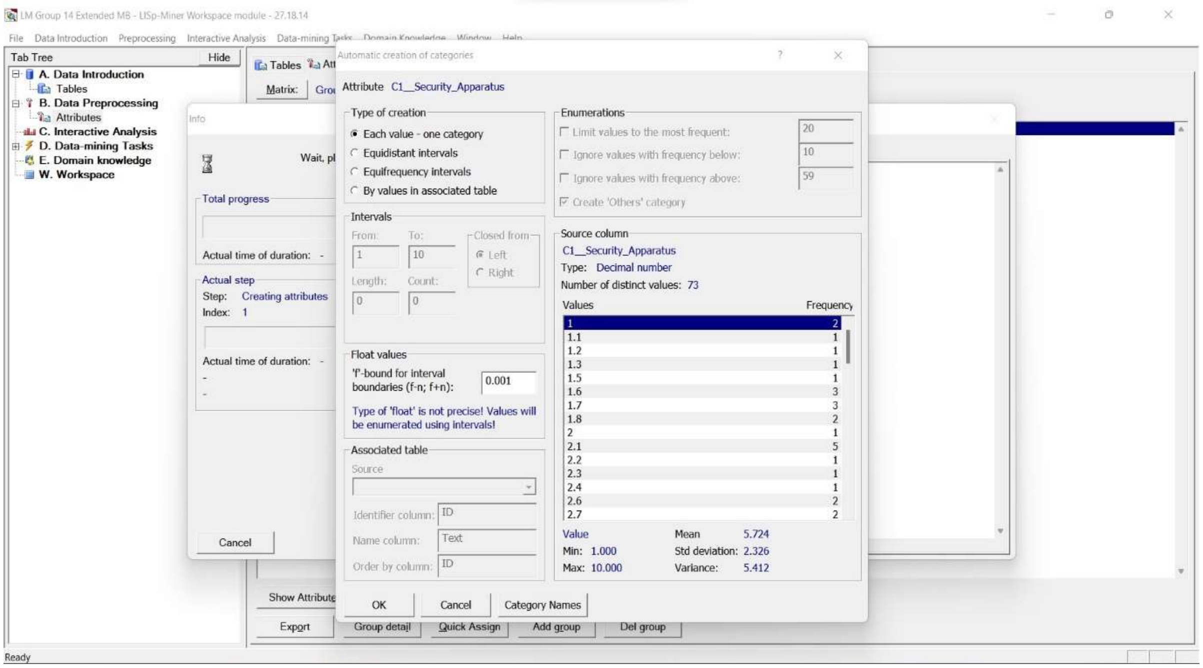


Figure 9.5: Changing the float value to 0.001 from 0.01.

All the attributes are now added except ID.

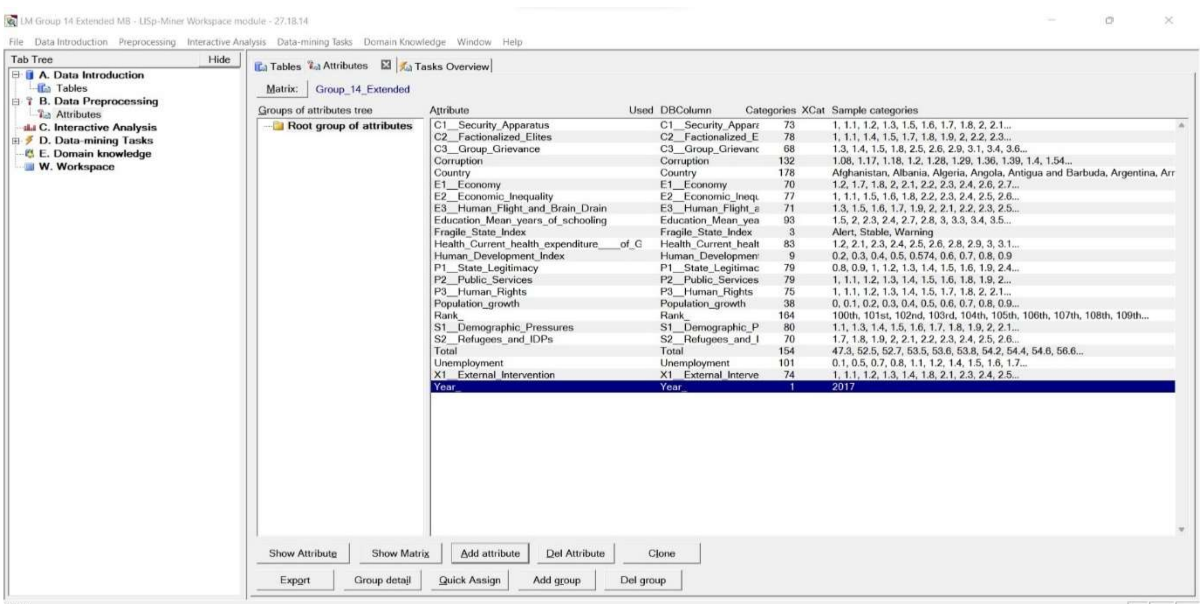


Figure 9.6: Selected Attributes.

Step 5: Click on Tasks Overview and add any task.

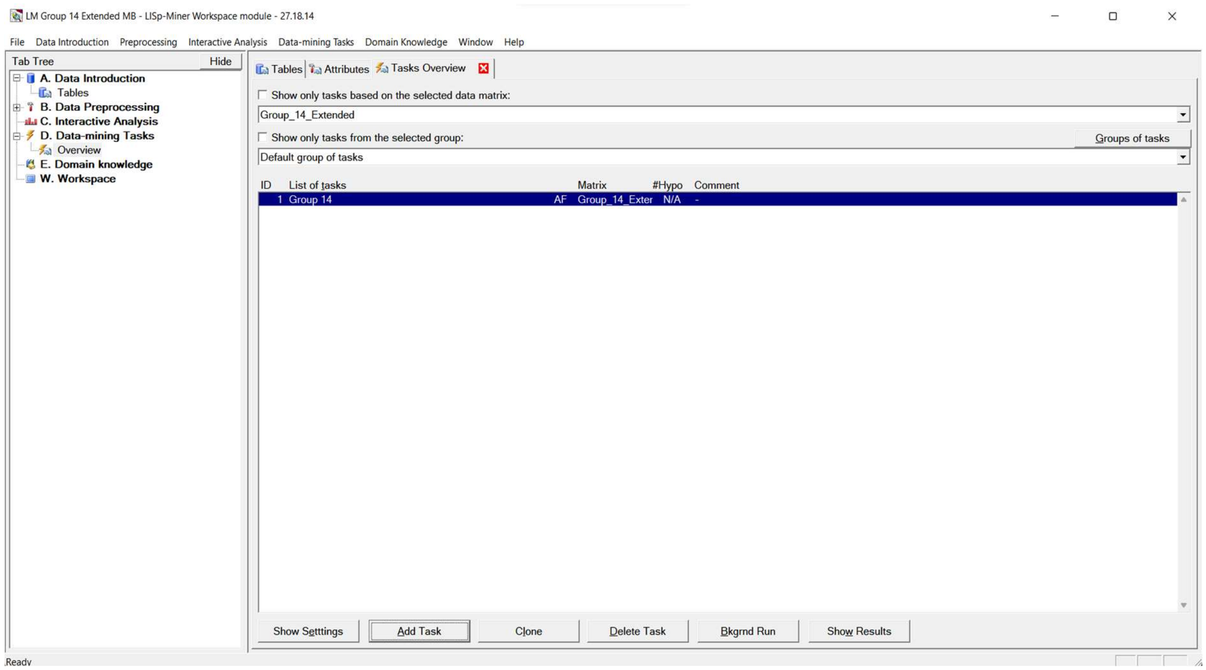


Figure 9.7: Adding the task in task overview.

Step 6: After adding the required attributes from the dataset, a new task is initialized. To perform Action rules tasks, we select Ac4ft-Miner which is an action rules miner. Now, select Ac4ft-Miner in Tasks Tab.

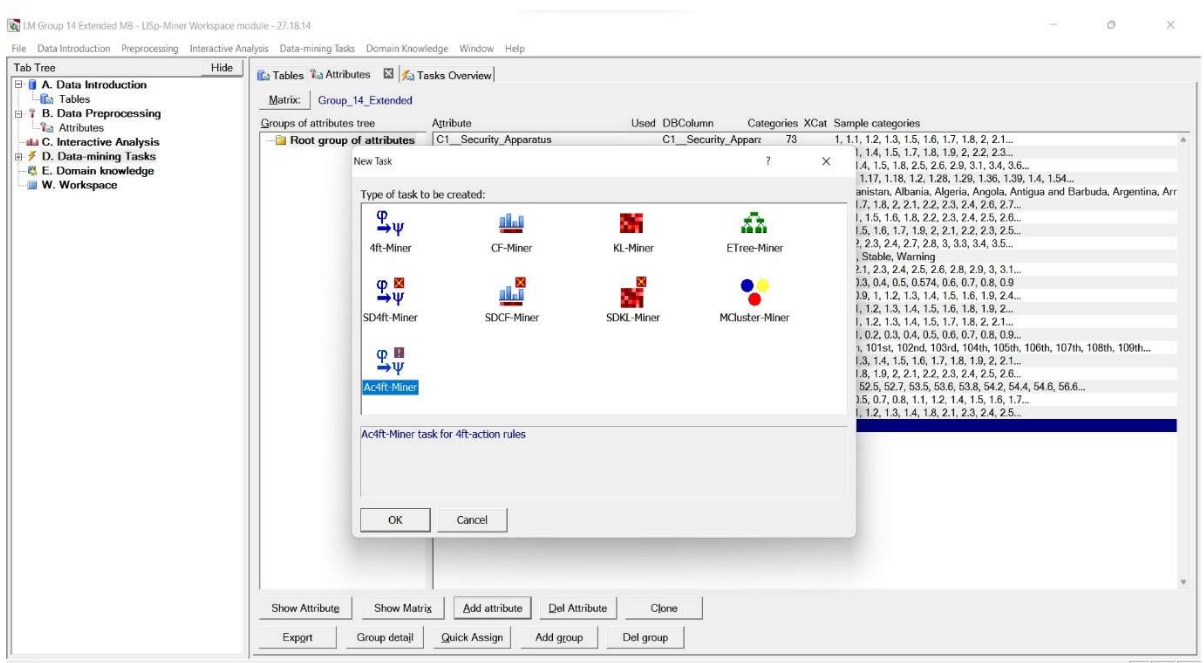


Figure 9.8: Selecting Ac4ft-Miner task to generate action rules.

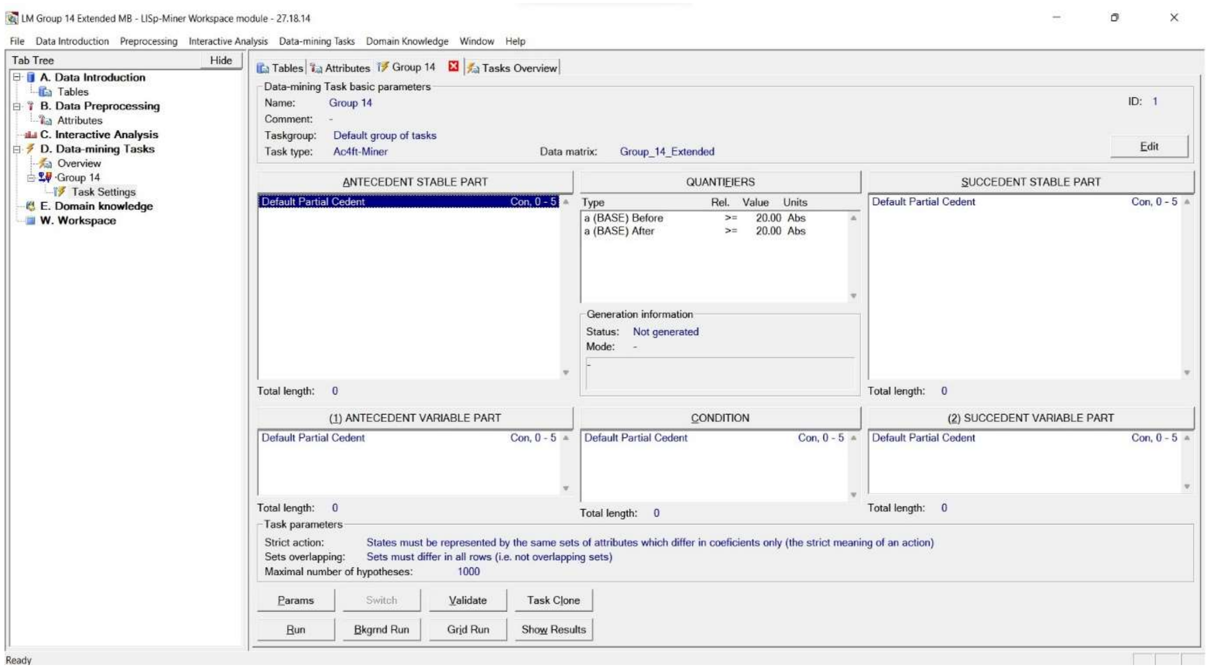


Figure 9.9: Parameter Settings.

Step 7: Selected Country, Rank, year attribute in the Antecedent stable part, Education_Mean_years_of_schooling and Unemployment in the Antecedent variable part. Selected Fragile State Index in the Succedent Variable part with the one category.

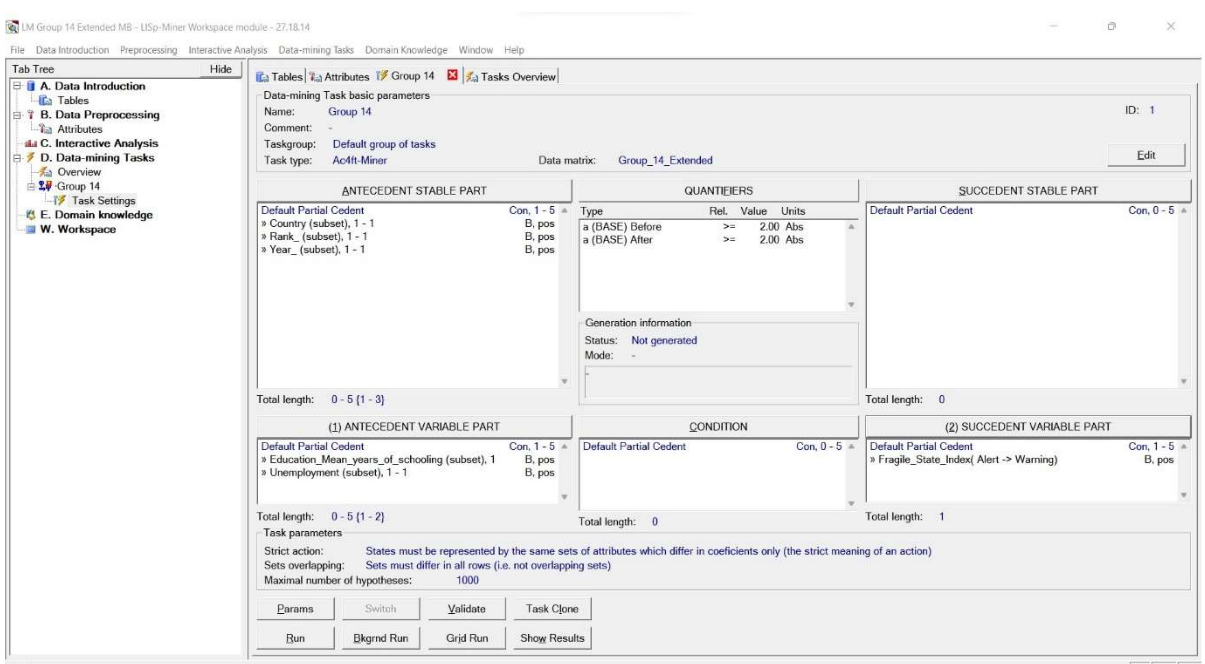


Figure 9.10: Added the attributes in Antecedent Stable and Variable Parts and Succedent Variable.

The coefficient type is one category, and the statebefore and stateafter attributes are set to Alert to Warning, implying that action rules will be developed for countries in the alert stage to move into the warning stage. A threshold of 2 in before and after value in the Quantifiers.

Step 8: Click Run

Output:

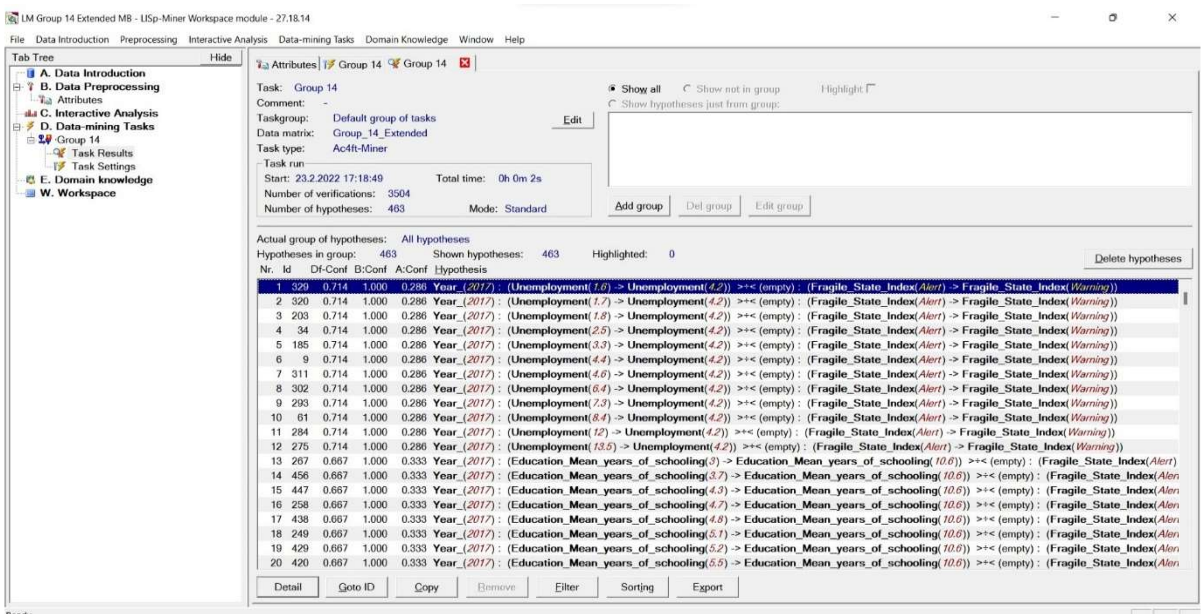


Figure 9.11: Task output which contains action rules.

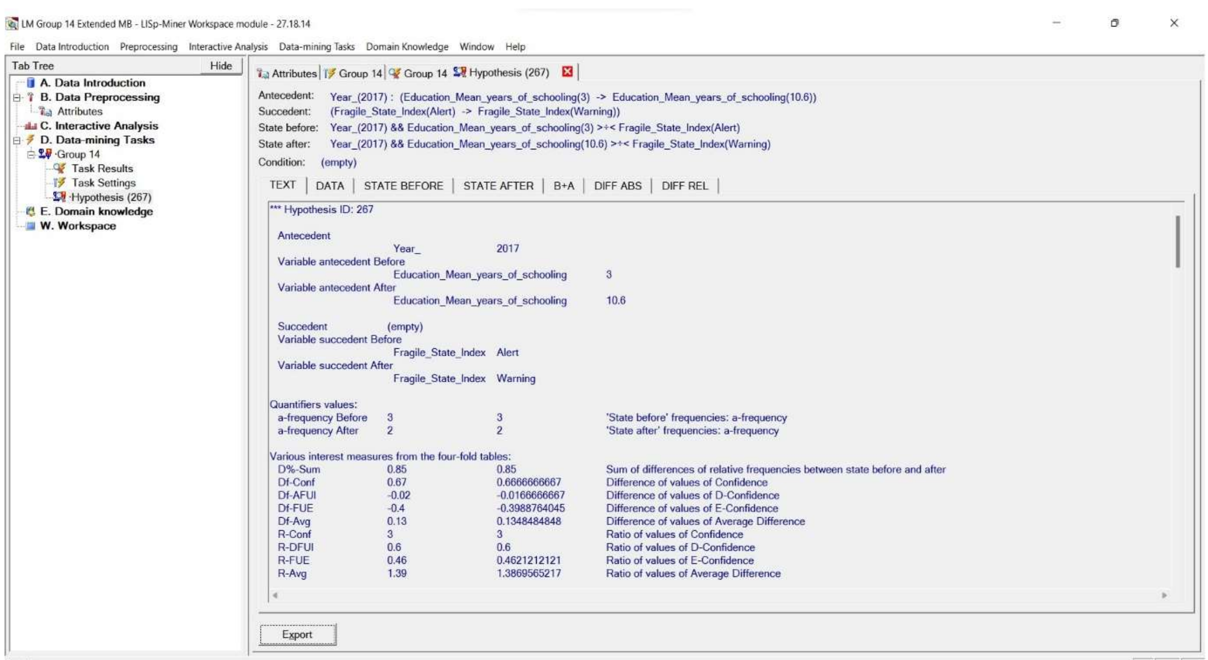


Figure 9.12: Task Hypothesis.

VISUALIZATION:

Action Rules-State Before:

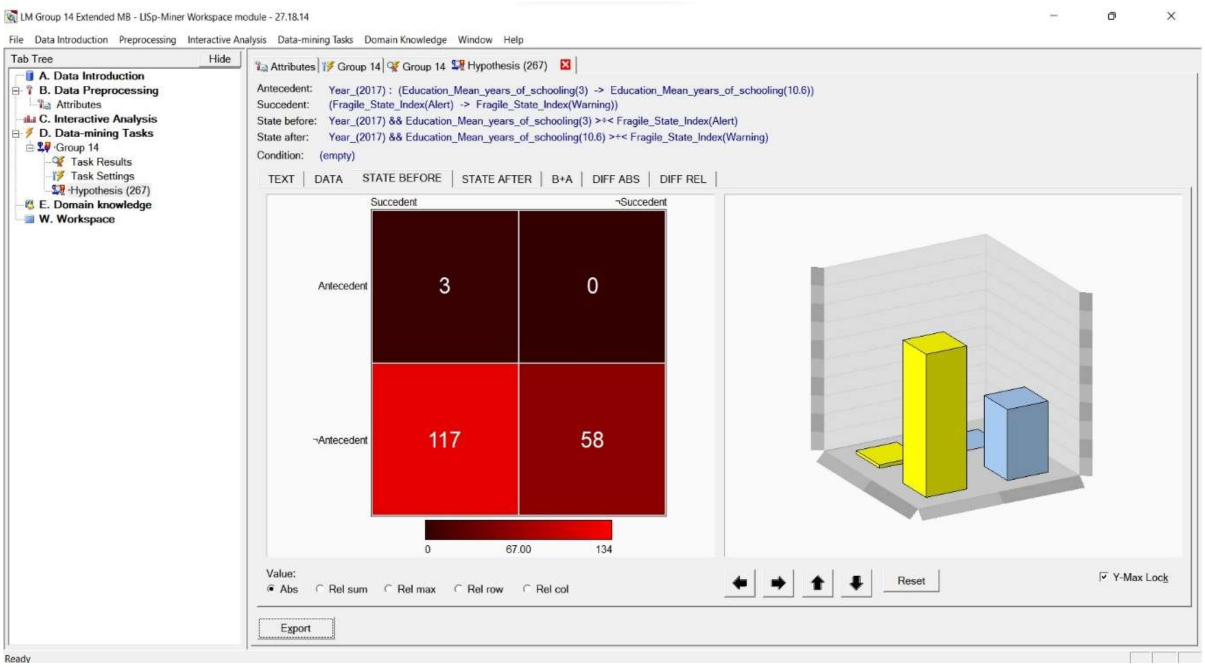


Figure 9.13: Hypothesis for State Before.

Action Rules-State After:

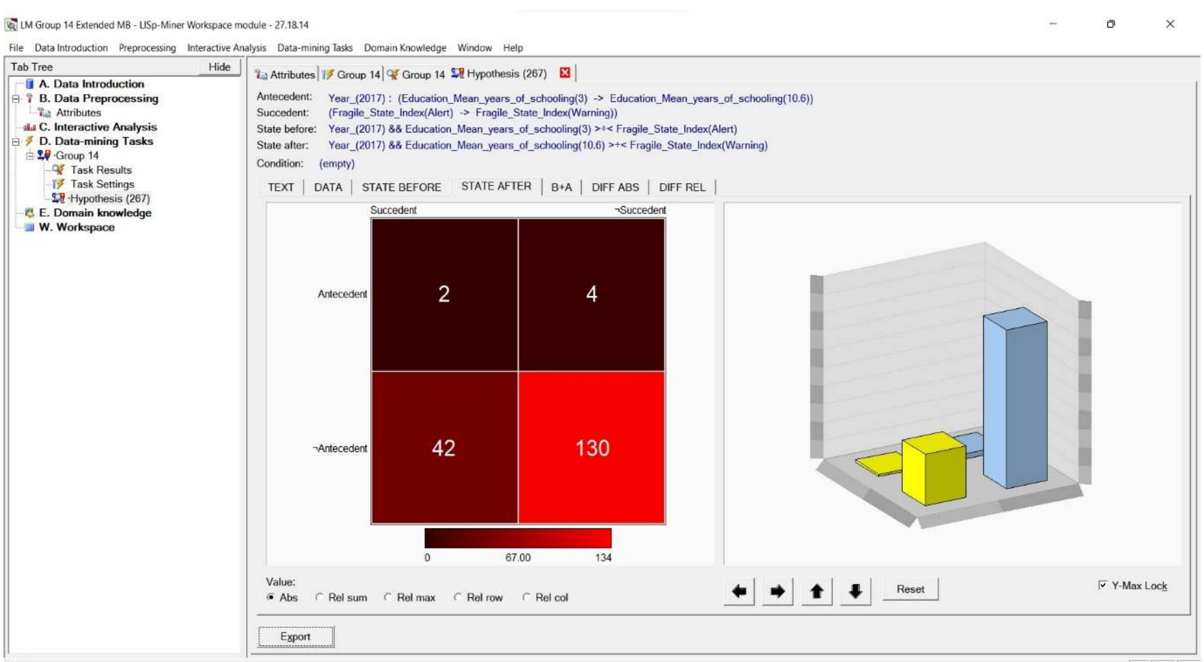


Figure 9.14: Hypothesis for State After.

Action Rules- B+A:

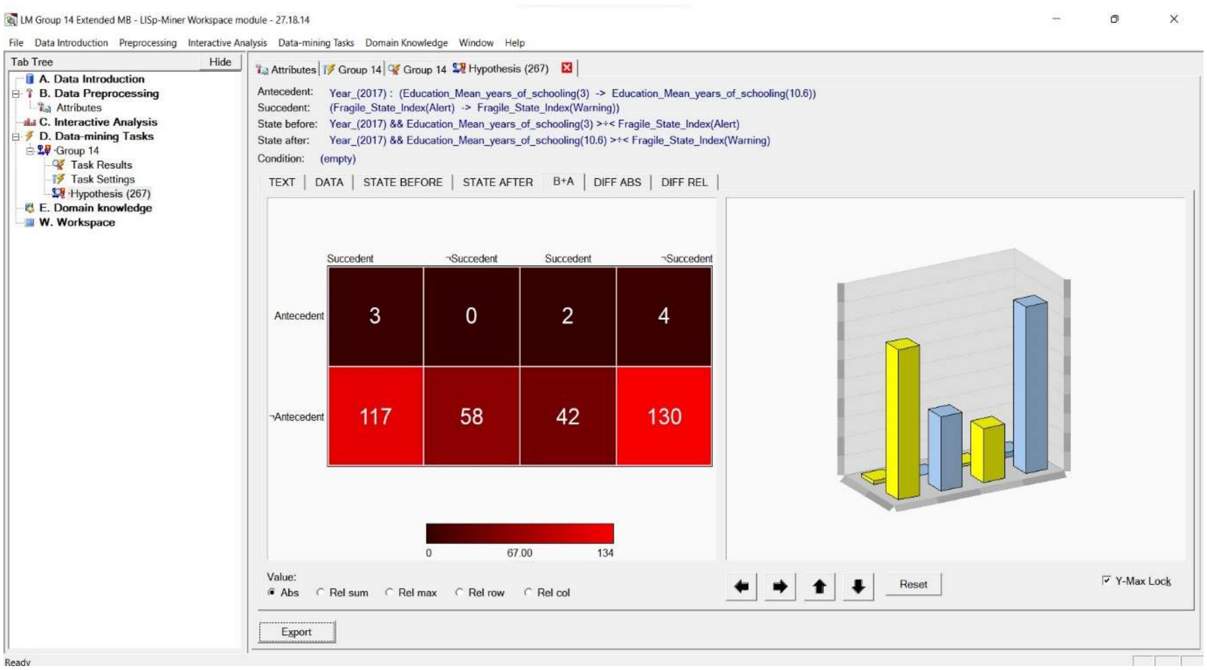


Figure 9.15: Hypothesis for B+A.

Action Rules DIFF ABS:

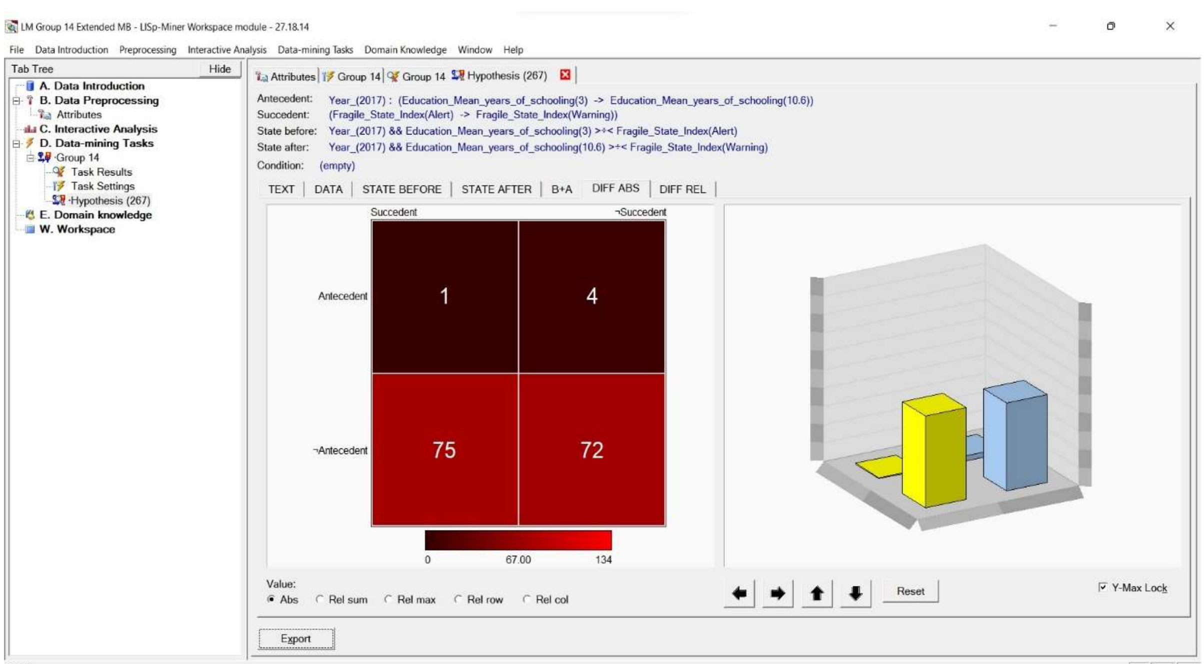


Figure 9.16: Hypothesis for DIFF ABS.

Action Rules- DIFF REL:

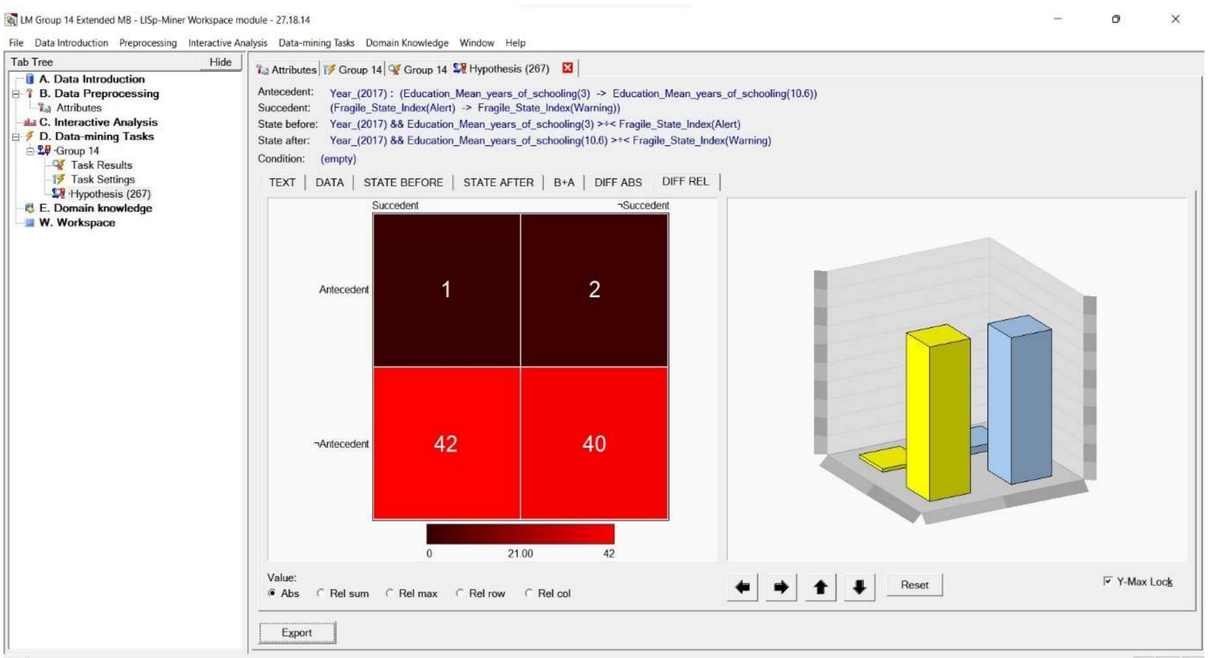


Figure 9.17: Hypothesis for DIFF REL.

10. FRAGILE ATTRIBUTES USED

1. **Education (Mean Years of schooling (Years)):** Average number of years of education gained by persons aged 25 and up, translated from education achievement levels using the official duration of each level. Increase the number of young people with necessary skills, including technical and vocational skills, for employment, good jobs, and entrepreneurship by 2030.

2. **Unemployment:** Percentage of the labor force population aged 15 and up who are not in paid employment or self-employment but are available for work and have taken actions to find paid employment or self-employment. Achieve full and productive employment and decent work for all women and men, including young people and those with disabilities, by 2030, as well as equal compensation for equal effort.

11. ACTION RULES AND INFERENCES

Stable Attributes: Country, Rank and Year

Flexible attributes: Education_Mean_years_of_schooling and Unemployment

Decision Attribute: Fragile State Index

Action Rules: (Boundary values)

1. (Education_Mean_years_of_schooling(3) -> Education_Mean_years_of_schooling(10.6)) >÷< (Fragile_State_Index(Alert) -> Fragile_State_Index(Warning))
2. (Education_Mean_years_of_schooling(8) -> Education_Mean_years_of_schooling(10.6)) >÷< (Fragile_State_Index(Alert) -> Fragile_State_Index(Warning))
3. (Unemployment(1.6) -> Unemployment(4.2)) >÷< (Fragile_State_Index(Alert) -> Fragile_State_Index(Warning))
4. (Unemployment(1.8) -> Unemployment(4.2)) >÷< (Fragile_State_Index(Alert) -> Fragile_State_Index(Warning))

The extracted rules are included in the text file below.

Group 14.txt

12. CONCLUSION

We added six new features to the current FSI dataset for 2017, each of which influences a country's Fragile States Index rating. The data was preprocessed using cleaning and normalization methods. The data was then discretized and categorized using the program WEKA. Classification Rules were discovered using WEKA. Following that, Lisp Miner was used to create Action rules. The Action rules that are developed might be considered as preventative measures that a country can take to improve its position. Screenshots of all the process is attached throughout. Also, we have attached all the results in the zip files for verification.

13. REFERENCES

- https://fragilestatesindex.org/wp-content/uploads/data/fsi-2017.xlsx
- https://en.wikipedia.org/wiki/Fragile_States_Index
- https://lispminer.vse.cz/demonstration/index.html

- https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/RandomForest.html
- http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayes.html
- https://weka.sourceforge.io/doc.dev/weka/classifiers/meta/RandomizableFilteredClassifier.html
- https://hdr.undp.org/en/indicators/103006#
- https://hdr.undp.org/en/indicators/140606#
- https://hdr.undp.org/en/indicators/137506
- https://databank.worldbank.org/
- https://www.transparency.org/en/cpi/2017
- https://hdr.undp.org/en/indicators/137506