

Analysis of Fragile States Data

**Department of Computer Science
University of North Carolina Charlotte,
NC, USA**

Instructor Dr. Zbigniew W. Ras

Team Members:

Dhananjay Arora (801077164)

Sumit Kawale (801135233)

Neelesh Mungoli (801083252)

Apnav Poptani (801137923)

INDEX

| | |
|---|----|
| • Project Description and Requirements..... | 3 |
| • Objective | |
| • Fragile State Index Description | |
| • Attributes | |
| • Extended Features and its motivation..... | 4 |
| • Source of the Dataset..... | 4 |
| • Data Extraction and Preprocessing..... | 5 |
| • Lisp Miner..... | 5 |
| • Data Discretization..... | 5 |
| • Classification using WEKA (screenshots included)..... | 7 |
| • Original Dataset | |
| • Extended (additional features) Dataset | |
| • Classification Accuracy Comparison Matrix..... | 11 |
| • Generation of Action Rules Using Lisp Miner..... | 11 |
| • Lisp Miner Screenshots..... | 12 |
| • Action rules and Interpretations..... | 13 |
| • Conclusions..... | 17 |
| • References..... | 17 |

Objective

The purpose of this project is to perform action rule mining for the FSI (Fragile State Index) dataset. We also need to add 6 new features to this dataset to do further analysis to suggest what changes are required in these features to lower the FSI. We can decide how one nation can move from a warning to a stable state using Action-Rules. The project shows the significance of each dataset attribute in determining the country's fragile state index. Discretization and identification of data are also carried out to analyze it better.

FSI Description

The Fragile State Index is an annual report that aims to assess states' vulnerability to conflict or collapse, ranking all sovereign states with membership in the United Nations that has a considerable amount of data to analyze. This ranking is based on the sum of the scores for the 12 indicators. Each indicator is scored on a scale of 0 to 10 with 10 being the highest intensity i.e., least stable and 0 being the lowest intensity i.e., most stable.

Following range is used to determine the fragile state of the country

- Alert - 90.0 to 120.0
- Warning - 60.0 to 89.9
- Stable - 30.0 to 59.9
- Sustainable - 0.0 to 29.9

Attributes Description

The following attributes are used to measure the condition of a state at any given time.

- Factionalized Elites
- Security Apparatus
- Group Grievance
- Economic Decline and Property
- Uneven Economic Development
- Human Flight and Brain Drain
- State Legitimacy
- Public Services
- Human Right and Rule of Law
- Demographic Pressures
- Refugees and Internally Displaced Persons
- External Intervention

Extended Features

In this report, along with 12 indicators of FSI, we are adding 6 new features to analyze the action rules. Following are the 6 new features and justifications for choosing these features:

- **Life Expectancy at Birth:** The life expectancy at birth refers to the average number of years a newborn is expected to live if mortality pattern at the time its birth remains constant in the future.
- **Women peace and security Index:** The women, peace, and security index offers a simple and transparent measure that captures women's autonomy and empowerment at home, in the community and in the society.
- **Military expenditure (% of Economy):** Military expenditure data from SIPRI are derived from NATO definition, which includes all current and capital expenditures on the armed forces, including peacekeeping forces, defense ministries, paramilitary forces.
- **Unemployment Rate:** Unemployment rate indicator refers to the share of the labor force that is without work but available for and seeking employment
- **Prevalence of Undernourishment:** This indicator refers to the share of the population which has a less caloric intake that is not sufficient to meet the necessary energy requirements for a given population. It is a leading risk factor for deaths and other health-related issues.
- **Death Rate from Obesity:** Obesity can be measured with the help of BMI, Body Mass Index scale. This scale has values that classify a person as underweight, healthy, overweight, or obese. People having BMI above 30.0 are considered as obese. It is one of the reasons for premature deaths. It causes problems such as heart disease, stroke, diabetes, and various types of cancer.

Motivation to select these features:

We randomly picked 10 features from World Bank's website 2017 data and applied different classifiers to it and selected the best 6 out of them based on whether the accuracy goes up or down based on the features chosen.

Source of the Dataset

The data for 6 newly added attributes has been taken from the below sources:

1. Life Expectancy at Birth: <https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2018&start=2016>
2. Women peace and security Index: <https://giwps.georgetown.edu/the-index/>
3. Military expenditure (% of Economy): <https://ourworldindata.org/military-spending>
4. Unemployment Rate: <https://ourworldindata.org/grapher/unemployment-rate>
5. Prevalence of Undernourishment: <https://ourworldindata.org/hunger-and-undernourishment>
6. Death Rate from Obesity: <https://ourworldindata.org/obesity>

Data Extraction and Preprocessing

- Data was extracted from the FSI website and the World Bank website.
- Some preprocessing was done in addition to it like dealing with the missing values(made as blank instead of NA), merging of data into one comma-separated values(CSV) file using 'vlookup' feature of Excel.
- After that, we performed data discretization and data classification.
- Action rules were extracted later on using LispMiner.

Lisp Miner

Lisp Miner is a tool that was developed to solve some of the problems associated with data mining. Lisp Miner has enabled the users to create new procedures and added the frameworks that were necessary for it to maintain compatibility among other modules. Lisp miner was developed at the University of Economics Prague. The main procedures that the system brought to the users were implementing a new module, giving the developer interfaces for Maintaining the cross-compatibility among the different types of modules.

The data analyzed is represented using bit strings that directly results in very fast data mining. The lucid architecture of the system enables the system to be customized for medicine, and finance. The main advantage of the system is to be able to cater to the needs of different professions and the Lisp miner system provides integration into large systems with relative ease.

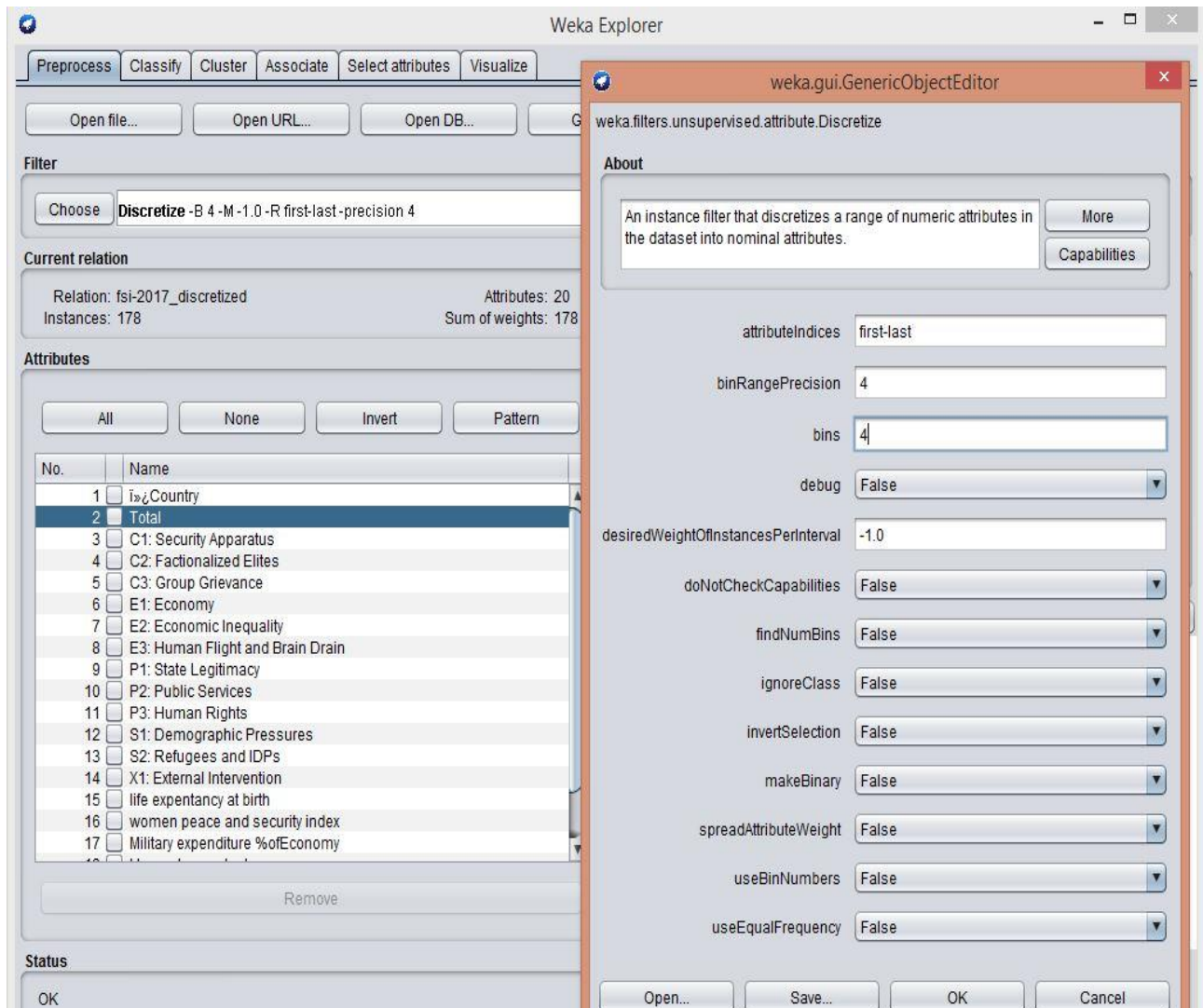
Data Discretization (Weka)

The tool called WEKA is used in this project for the classification of our data.

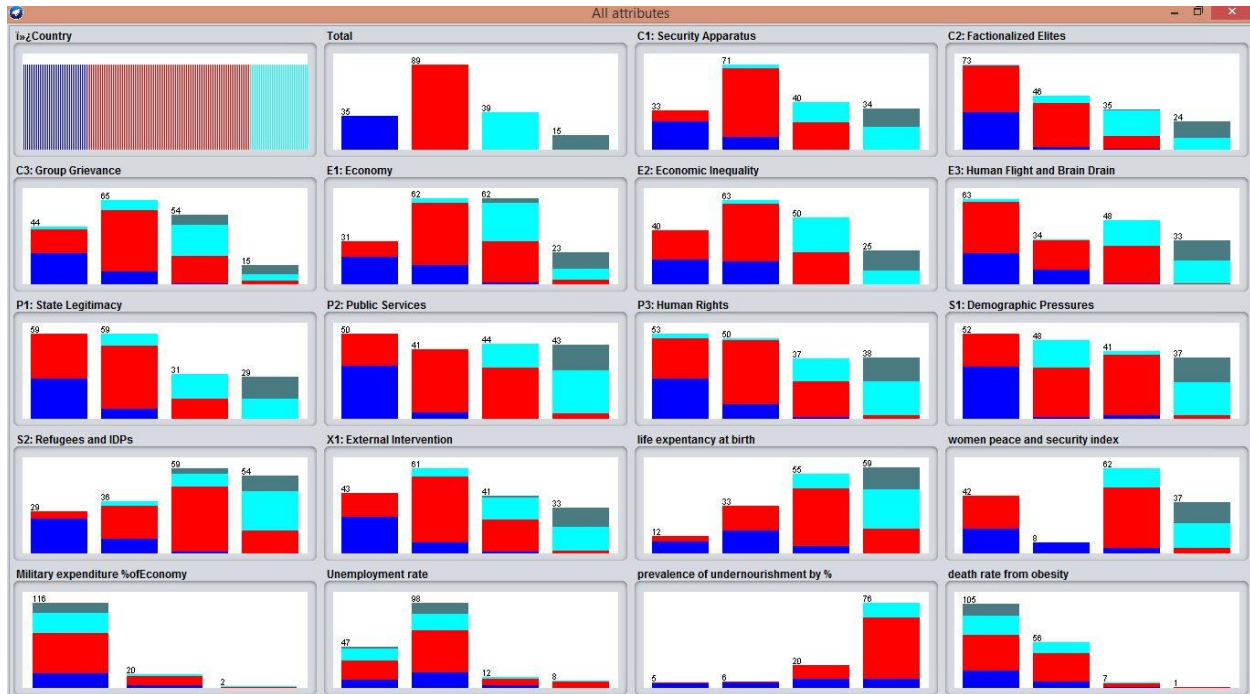
The discretization is a process that consists of converting or partitioning continuous attributes, features, or variables to discrete or nominal values. The decision attribute TOTAL has continuous values throughout the dataset ranging from 0 to 120. We have used discretization to replace these numeric values with the following categories.

- Alert: 90-120
- Warning: 60-90
- Stable: 30-60
- Sustainable: 0-30

To perform effective classification, we have discretized other attributes from numeric values to nominal values.



Following chart shows the Discretized data:



Data Classification (WEKA)

In the data classification process, the data is organized by relevant categories so that it may be used and protected more efficiently. In this project, the following classification algorithms were used for data classification.

1. **Logistic Model Trees (LMT):** The LMT is a supervised training classification algorithm. It is made by integrating standard decision tree induction and linear logistic regression algorithms in a single tree. In LMT, cross-validation is used to find several LogitBoost(algorithm) iterations that do not overfit the training data.
2. **Random Forest:** The Random Forest classification is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.
3. **Simple Logistic:** The logistic classification model is a binary classification model in which the conditional probability of one of the two possible realizations of the output variable is assumed to be equal to a linear combination of the input variables.

Weka Screenshots

1) Logistic Model Trees

Before adding new features:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'Simple Logistic - I 0 - M 500 - H 50 - W 0.0'. The 'Test options' are set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

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

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 151 | 84.8315 % |
| Incorrectly Classified Instances | 27 | 15.1685 % |
| Kappa statistic | 0.7664 | |
| Mean absolute error | 0.0862 | |
| Root mean squared error | 0.2266 | |
| Relative absolute error | 26.1851 % | |
| Root relative squared error | 55.9332 % | |
| Total Number of Instances | 178 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| Alert | 0.743 | 0.035 | 0.839 | 0.743 | 0.788 | 0.742 | 0.980 | 0.931 | Alert |
| Warning | 0.899 | 0.157 | 0.851 | 0.899 | 0.874 | 0.743 | 0.959 | 0.960 | Warning |
| Stable | 0.795 | 0.036 | 0.861 | 0.795 | 0.827 | 0.782 | 0.969 | 0.921 | Stable |
| Sustainable | 0.933 | 0.018 | 0.824 | 0.933 | 0.875 | 0.865 | 0.992 | 0.901 | Sustainable |
| Weighted Avg. | 0.848 | 0.095 | 0.849 | 0.848 | 0.847 | 0.761 | 0.968 | 0.941 | |

=== Confusion Matrix ===

| | a | b | c | d | <-- classified as |
|----|----|----|----|---|-------------------|
| 26 | 9 | 0 | 0 | 1 | a = Alert |
| 5 | 80 | 4 | 0 | 1 | b = Warning |
| 0 | 5 | 31 | 3 | 1 | c = Stable |
| 0 | 0 | 1 | 14 | 1 | d = Sustainable |

After adding new features:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'Simple Logistic - I 0 - M 500 - H 50 - W 0.0'. The 'Test options' are set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

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

| Metric | Value | Percentage |
|----------------------------------|-----------|------------|
| Correctly Classified Instances | 154 | 86.5169 % |
| Incorrectly Classified Instances | 24 | 13.4831 % |
| Kappa statistic | 0.7942 | |
| Mean absolute error | 0.0834 | |
| Root mean squared error | 0.2201 | |
| Relative absolute error | 25.3319 % | |
| Root relative squared error | 54.3174 % | |
| Total Number of Instances | 178 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| Alert | 0.771 | 0.042 | 0.818 | 0.771 | 0.794 | 0.746 | 0.966 | 0.914 | Alert |
| Warning | 0.888 | 0.124 | 0.878 | 0.888 | 0.883 | 0.764 | 0.957 | 0.949 | Warning |
| Stable | 0.872 | 0.036 | 0.872 | 0.872 | 0.872 | 0.836 | 0.973 | 0.938 | Stable |
| Sustainable | 0.933 | 0.012 | 0.875 | 0.933 | 0.903 | 0.895 | 0.993 | 0.904 | Sustainable |
| Weighted Avg. | 0.865 | 0.079 | 0.865 | 0.865 | 0.865 | 0.787 | 0.965 | 0.936 | |

=== Confusion Matrix ===

| | a | b | c | d | <-- classified as |
|----|----|----|----|---|-------------------|
| 27 | 8 | 0 | 0 | 1 | a = Alert |
| 6 | 79 | 4 | 0 | 1 | b = Warning |
| 0 | 3 | 34 | 2 | 1 | c = Stable |
| 0 | 0 | 1 | 14 | 1 | d = Sustainable |

2) Random Forest

Before adding new features:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'Simple Logistic - I 0 - M 500 - H 50 - W 0.0'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

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

| Metric | Value |
|----------------------------------|----------|
| Correctly Classified Instances | 154 |
| Incorrectly Classified Instances | 24 |
| Kappa statistic | 0.7851 |
| Mean absolute error | 0.2113 |
| Root mean squared error | 0.2753 |
| Relative absolute error | 64.145 % |
| Root relative squared error | 67.931 % |
| Total Number of Instances | 178 |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| Weighted Avg. | 0.686 | 0.014 | 0.923 | 0.686 | 0.787 | 0.756 | 0.986 | 0.950 | Alert |
| | 0.978 | 0.213 | 0.821 | 0.978 | 0.892 | 0.778 | 0.974 | 0.974 | Warning |
| | 0.769 | 0.014 | 0.938 | 0.769 | 0.845 | 0.813 | 0.989 | 0.967 | Stable |
| | 0.867 | 0.006 | 0.929 | 0.867 | 0.897 | 0.888 | 0.999 | 0.991 | Sustainable |

=== Confusion Matrix ===

| a | b | c | d | <-- classified as |
|----|----|----|----|-------------------|
| 24 | 11 | 0 | 0 | a = Alert |
| 2 | 87 | 0 | 0 | b = Warning |
| 0 | 8 | 30 | 1 | c = Stable |
| 0 | 0 | 2 | 13 | d = Sustainable |

After adding new features:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'Simple Logistic - I 0 - M 500 - H 50 - W 0.0'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

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

| Metric | Value |
|----------------------------------|-----------|
| Correctly Classified Instances | 158 |
| Incorrectly Classified Instances | 20 |
| Kappa statistic | 0.8219 |
| Mean absolute error | 0.1959 |
| Root mean squared error | 0.2601 |
| Relative absolute error | 59.4914 % |
| Root relative squared error | 64.1881 % |
| Total Number of Instances | 178 |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| Weighted Avg. | 0.743 | 0.007 | 0.963 | 0.743 | 0.839 | 0.815 | 0.989 | 0.957 | Alert |
| | 0.989 | 0.180 | 0.846 | 0.989 | 0.912 | 0.821 | 0.978 | 0.979 | Warning |
| | 0.795 | 0.014 | 0.939 | 0.795 | 0.861 | 0.831 | 0.988 | 0.965 | Stable |
| | 0.867 | 0.006 | 0.929 | 0.867 | 0.897 | 0.888 | 0.999 | 0.992 | Sustainable |

=== Confusion Matrix ===

| a | b | c | d | <-- classified as |
|----|----|----|----|-------------------|
| 26 | 9 | 0 | 0 | a = Alert |
| 1 | 88 | 0 | 0 | b = Warning |
| 0 | 7 | 31 | 1 | c = Stable |
| 0 | 0 | 2 | 13 | d = Sustainable |

3) Simple Logistic

Before adding new features:

The Weka Explorer window displays the results of a Simple Logistic classifier. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section shows the following summary statistics:

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      151          84.8315 %
Incorrectly Classified Instances    27           15.1685 %
Kappa statistic                    0.7664
Mean absolute error                 0.0862
Root mean squared error             0.2266
Relative absolute error             26.1851 %
Root relative squared error         55.9332 %
Total Number of Instances          178
  
```

The 'Detailed Accuracy By Class' table is as follows:

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| | 0.743 | 0.035 | 0.839 | 0.743 | 0.788 | 0.742 | 0.980 | 0.931 | Alert |
| | 0.899 | 0.157 | 0.851 | 0.899 | 0.874 | 0.743 | 0.959 | 0.960 | Warning |
| | 0.795 | 0.036 | 0.861 | 0.795 | 0.827 | 0.782 | 0.969 | 0.921 | Stable |
| | 0.933 | 0.018 | 0.824 | 0.933 | 0.875 | 0.865 | 0.992 | 0.901 | Sustainable |
| Weighted Avg. | 0.848 | 0.095 | 0.849 | 0.848 | 0.847 | 0.761 | 0.968 | 0.941 | |

The 'Confusion Matrix' is shown below:

```

=== Confusion Matrix ===
 a b c d <-- classified as
26 9 0 0 | a = Alert
 5 80 4 0 | b = Warning
 0 5 31 3 | c = Stable
 0 0 1 14 | d = Sustainable
  
```

After adding new features:

The Weka Explorer window displays the results of a Simple Logistic classifier after adding new features. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section shows the following summary statistics:

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      154          86.5169 %
Incorrectly Classified Instances    24           13.4831 %
Kappa statistic                    0.7942
Mean absolute error                 0.0834
Root mean squared error             0.2201
Relative absolute error             25.3319 %
Root relative squared error         54.3174 %
Total Number of Instances          178
  
```

The 'Detailed Accuracy By Class' table is as follows:

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------------|
| | 0.771 | 0.042 | 0.818 | 0.771 | 0.794 | 0.746 | 0.966 | 0.914 | Alert |
| | 0.888 | 0.124 | 0.878 | 0.888 | 0.883 | 0.764 | 0.957 | 0.949 | Warning |
| | 0.872 | 0.036 | 0.872 | 0.872 | 0.872 | 0.836 | 0.973 | 0.938 | Stable |
| | 0.933 | 0.012 | 0.875 | 0.933 | 0.903 | 0.895 | 0.993 | 0.904 | Sustainable |
| Weighted Avg. | 0.865 | 0.079 | 0.865 | 0.865 | 0.865 | 0.787 | 0.965 | 0.936 | |

The 'Confusion Matrix' is shown below:

```

=== Confusion Matrix ===
 a b c d <-- classified as
27 8 0 0 | a = Alert
 6 79 4 0 | b = Warning
 0 3 34 2 | c = Stable
 0 0 1 14 | d = Sustainable
  
```

The following **classification matrix** contains correctly classified data which was built using the above results:

| <i>Classification Algorithms</i> | <i>Before Adding Attributes</i> | <i>After Adding Attributes</i> |
|---|--|---------------------------------------|
| LMT | 84.83% | 86.52% |
| Random Forest | 86.51% | 88.76% |
| Simple Logistic | 84.83% | 86.52% |

Generation of Action Rules Using LISP Miner

We have grouped our attributes into 3 categories

- Stable
 - Country
- Flexible
 - Security apparatus
 - Economy
 - Life expectancy at birth
 - Human rights
 - Demographic pressures
- Decision
 - Total

Antecedent Stable Part:

We took attributes from the stable category: Country

Antecedent Variable Part:

We took attributes from the flexible category: Security apparatus, Economy, Life expectancy at birth, Human rights, and Demographic pressures.

Succedent Variable Part:

We assigned a decision variable to this set.

Attribute: Total

Coefficient Type: One category

Quantifiers:

a(Base) Before: 10

a(Base) After: 10

Extracting Action Rules:

Create a task and set decision from the Alert to Warning and Add Antecedent Stable and Antecedent Variable Part

The screenshot displays the LISP-Miner Workspace module interface. The left sidebar shows a tree view with categories: A. Data Introduction, B. Data Preprocessing, C. Interactive Analysis, D. Data-mining Tasks, E. Domain knowledge, and W. Workspace. Under 'D. Data-mining Tasks', 'Task Settings' is selected, showing 'Hypothesis (14)'. The main window is titled 'LM fsi-2017_discretized MB - LISP-Miner Workspace module - 27.18.07'. It contains a 'Data-mining Task basic parameters' section with fields for Name ('?- Task'), Comment, Taskgroup ('Default group of tasks'), Task type ('Ac4ft-Miner'), Data matrix ('fsi_2017_discretized'), and ID ('1'). Below this are three main sections: 'ANTECEDENT STABLE PART', 'QUANTIFIERS', and 'SUCCEDENT STABLE PART'. The 'ANTECEDENT STABLE PART' shows a list of items with coefficients and positions, including 'Default Partial Cedent' (Con, 0 - 5) and '» d_Country (subset), 1 - 1' (B, pos). The 'QUANTIFIERS' section has a table with columns 'Type', 'Rel.', 'Value', and 'Units', containing 'a (BASE) Before' (>=, 10.00, Abs) and 'a (BASE) After' (>=, 10.00, Abs). The 'SUCCEDENT STABLE PART' shows 'Default Partial Cedent' (Con, 0 - 5). Below these are three more sections: '(1) ANTECEDENT VARIABLE PART', 'CONDITION', and '(2) SUCCEDENT VARIABLE PART'. The '(1) ANTECEDENT VARIABLE PART' lists items like '» x_C1_Security_Apparatus_ (subse' (B, pos) and '» x_E1_Economy_ (subset), 1 - 1' (B, pos). The 'CONDITION' section shows 'Default Partial Cedent' (Con, 0 - 5). The '(2) SUCCEDENT VARIABLE PART' shows '» Total(Alert -> Warning)' (B, pos). At the bottom, there are buttons for 'Params', 'Switch', 'Validate', 'Task Clone', 'Run', 'Bkgnd Run', 'Grid Run', and 'Show Results'. The status bar at the bottom indicates 'Data...' and '100%'.

Tab Tree: Hide

File Data Introduction Preprocessing Interactive Analysis Data-mining Tasks Domain Knowledge Window Help

Tables Attributes Tasks Overview ?- Task ?- Task Hypothesis (14)

Data-mining Task basic parameters

Name: ?- Task ID: 1

Comment: -

Taskgroup: Default group of tasks

Task type: Ac4ft-Miner Data matrix: fsi_2017_discretized Edit

ANTECEDENT STABLE PART

Default Partial Cedent Con, 0 - 5

» d_Country (subset), 1 - 1 B, pos

QUANTIFIERS

| Type | Rel. | Value | Units |
|-----------------|------|-------|-------|
| a (BASE) Before | >= | 10.00 | Abs |
| a (BASE) After | >= | 10.00 | Abs |

Generation information

Status: Solved, 2 run(s)

Mode: Standard

SUCCEDENT STABLE PART

Default Partial Cedent Con, 0 - 5

Total length: 0 - 5 (0 - 1)

(1) ANTECEDENT VARIABLE PART

Default Partial Cedent Con, 0 - 5

» x_C1_Security_Apparatus_ (subse B, pos

» x_E1_Economy_ (subset), 1 - 1 B, pos

» X_life_expectancy_at_birth_ (subse B, pos

» x_P3_Human Rights (subset), 1 - B, pos

CONDITION

Default Partial Cedent Con, 0 - 5

(2) SUCCEDENT VARIABLE PART

Default Partial Cedent Con, 1 - 2

» Total(Alert -> Warning) B, pos

Total length: 0

Total length: 1

Task parameters

Strict action: States must be represented by the same sets of attributes which differ in coefficients only (the strict meaning of an action)

Sets overlapping: Sets must differ in all rows (i.e. not overlapping sets)

Maximal number of hypotheses: 1000

Params Switch Validate Task Clone

Run Bkgnd Run Grid Run Show Results

After running, we get the task result (Action Rules)

LM fsi-2017_discretized MB - LSP-Miner Workspace module - 27.18.07

File Data Introduction Preprocessing Interactive Analysis Data-mining Tasks Domain Knowledge Window Help

Tab Tree: A. Data Introduction, B. Data Preprocessing, C. Interactive Analysis, D. Data-mining Tasks, E. Domain knowledge, W. Workspace

Task: ? - Task
Comment: -
Taskgroup: Default group of tasks
Data matrix: fsi_2017_discretized
Task type: Ac4ft-Miner

Task run
Start: 29.4.2020 16:38:25
Total time: 0h 0m 0s
Number of verifications: 234
Number of hypotheses: 22
Mode: Standard

Actual group of hypotheses: All hypotheses
Hypotheses in group: 22
Shown hypotheses: 22
Highlighted: 0

| Nr. | ID | Df-Conf | B-Conf | A-Conf | Hypothesis |
|-----|----|---------|--------|--------|---|
| 1 | 14 | 0.291 | 0.846 | 0.556 | (empty): (X__life_expentancy_at_birth_('\(60.2049-68.1699]')) & x__P3__Human_Rights_('\(7.6-inf)')) >>> (empty): (X__life_expentancy_at_birth_('\(68.1699-76.1348]')) & x__P3__Human_Rights_('\(3.2-5.4]')) |
| 2 | 13 | 0.197 | 0.485 | 0.288 | (empty): (X__life_expentancy_at_birth_('\(60.2049-68.1699]')) >>> (empty): (X__life_expentancy_at_birth_('\(76.1348-inf)')) >>> (empty): (X__life_expentancy_at_birth_('\(68.1699-76.1348]')) & x__P3__Human_Rights_('\(3.2-5.4]')) |
| 3 | 4 | 0.172 | 0.941 | 0.769 | (empty): (x__C1__Security_Apparatus_('\(7.75-inf)')) & x__E1__Economy_('\(7.8-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 4 | 9 | 0.161 | 0.645 | 0.484 | (empty): (x__E1__Economy_('\(3.4-5.6]')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 5 | 2 | 0.152 | 0.727 | 0.575 | (empty): (x__C1__Security_Apparatus_('\(7.75-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 6 | 5 | 0.123 | 0.941 | 0.818 | (empty): (x__C1__Security_Apparatus_('\(7.75-inf)')) & x__E1__Economy_('\(3.25-5.5]')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 7 | 6 | 0.099 | 0.941 | 0.842 | (empty): (x__C1__Security_Apparatus_('\(7.75-inf)')) & x__E1__Economy_('\(5.6-7.8]')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 8 | 21 | -0.010 | 0.615 | 0.625 | (empty): (x__S1__Demographic_Pressures_('\(7.75-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 9 | 19 | -0.071 | 0.846 | 0.917 | (empty): (x__P3__Human_Rights_('\(7.6-inf)')) & x__S1__Demographic_Pressures_('\(7.75-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 10 | 3 | -0.076 | 0.727 | 0.803 | (empty): (x__C1__Security_Apparatus_('\(7.75-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 11 | 10 | -0.081 | 0.645 | 0.726 | (empty): (x__E1__Economy_('\(7.8-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 12 | 16 | -0.091 | 0.909 | 1.000 | (empty): (X__life_expentancy_at_birth_('\(60.2049-68.1699]')) & x__S1__Demographic_Pressures_('\(7.75-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 13 | 11 | -0.094 | 0.485 | 0.579 | (empty): (X__life_expentancy_at_birth_('\(60.2049-68.1699]')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 14 | 15 | -0.106 | 0.846 | 0.952 | (empty): (X__life_expentancy_at_birth_('\(60.2049-68.1699]')) & x__P3__Human_Rights_('\(7.6-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 15 | 17 | -0.123 | 0.472 | 0.595 | (empty): (x__P3__Human_Rights_('\(7.6-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 16 | 8 | -0.129 | 0.226 | 0.355 | (empty): (x__E1__Economy_('\(5.6-7.8]')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 17 | 20 | -0.154 | 0.846 | 1.000 | (empty): (x__P3__Human_Rights_('\(7.6-inf)')) & x__S1__Demographic_Pressures_('\(7.75-inf)')) >>> (empty): (Total(Alert) > Total(Warning)) |
| 18 | 12 | -0.242 | 0.485 | 0.727 | (empty): (X__life_expentancy_at_birth_('\(60.2049-68.1699]')) >>> (empty): (Total(Alert) > Total(Warning)) |

Buttons: Detail, Goto ID, Copy, Remove, Filter, Sorting, Export

Selecting Action Rules

1----(X__life_expentancy_at_birth_('\(60.2049-68.1699]')) & x__P3__Human_Rights_('\(7.6-inf)')) ----->

X__life_expentancy_at_birth_('\(68.1699-76.1348]')) & x__P3__Human_Rights_('\(3.2-5.4]'))

Converts Decision From:

Alert --> Warning

2----(X__life_expentancy_at_birth_('\(60.2049-68.1699]')) ----->

X__life_expentancy_at_birth_('\(76.1348-inf)'))

Converts Decision From:

Alert --> Warning

3----(x__C1__Security_Apparatus_('\(7.75-inf)')) & x__E1__Economy_('\(7.8-inf)')) ----->

x__C1__Security_Apparatus_('\(3.25-5.5]')) & x__E1__Economy_('\(5.6-7.8]'))

Converts Decision From:

Alert --> Warning

4----(x__E1__Economy_('\(7.8-inf)\')) -----> x__E1__Economy_('\(3.4-5.6]\'))

Converts Decision From:

Alert --> Warning

Interpretation of these rules:

We can change the FSI status of the country using above-extracted action rules.

For example:

Action Rule 1---- If the Life expectancy at birth and the Human Rights of a country is improved according to above-specified metrics then the FSI status of the country can be improved from Alert to Warning.

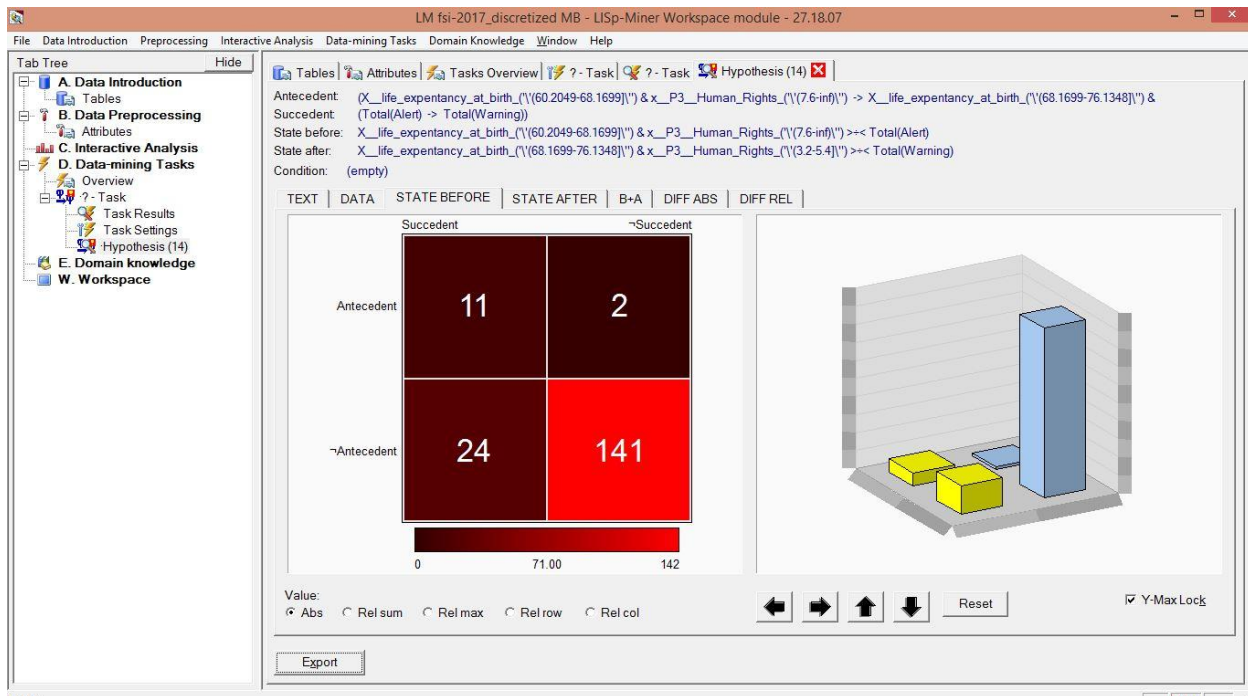
Action Rule 2--- If the Life Expectancy at Birth is improved according to the above-specified metrics then the FSI status of the country can be improved from Alert to Warning.

Action Rule 3--- If the Security Apparatus and the Economy of a country is improved according to above-specified metrics then the FSI status of the country can be improved from Alert to Warning.

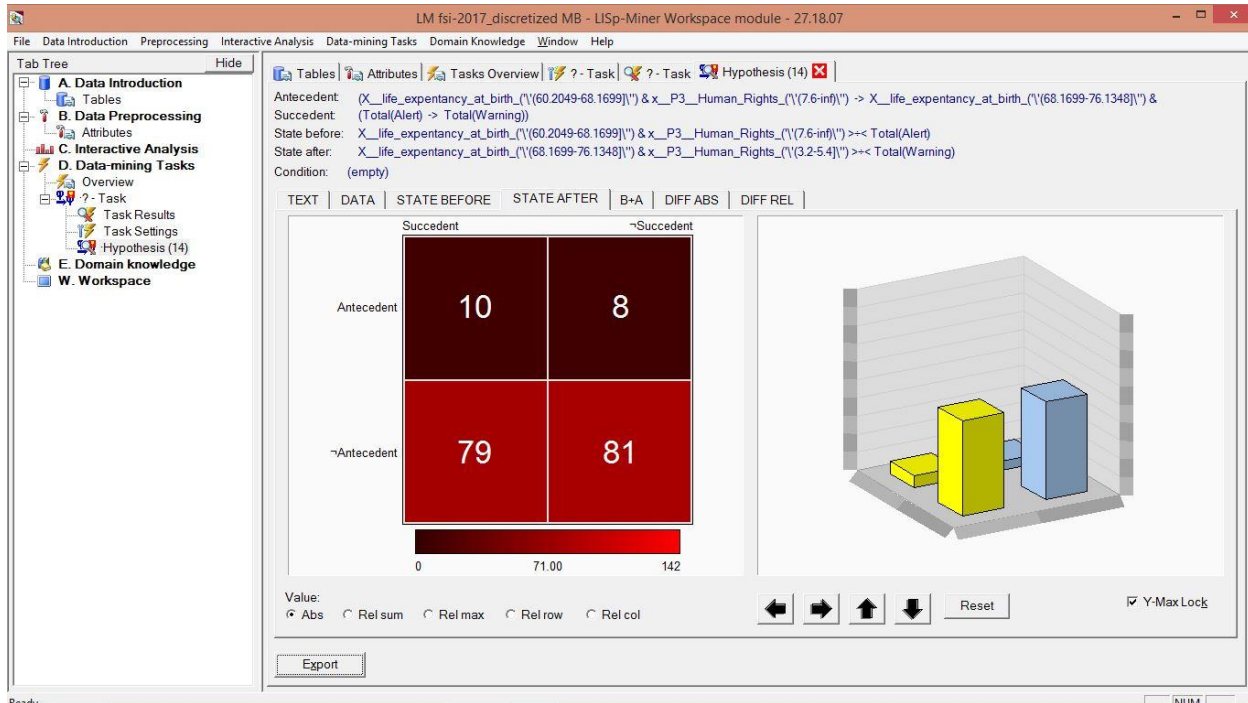
Action Rule 4--- If the Economy of a country is improved according to above-specified metrics then the FSI status of the country can be improved from Alert to Warning.

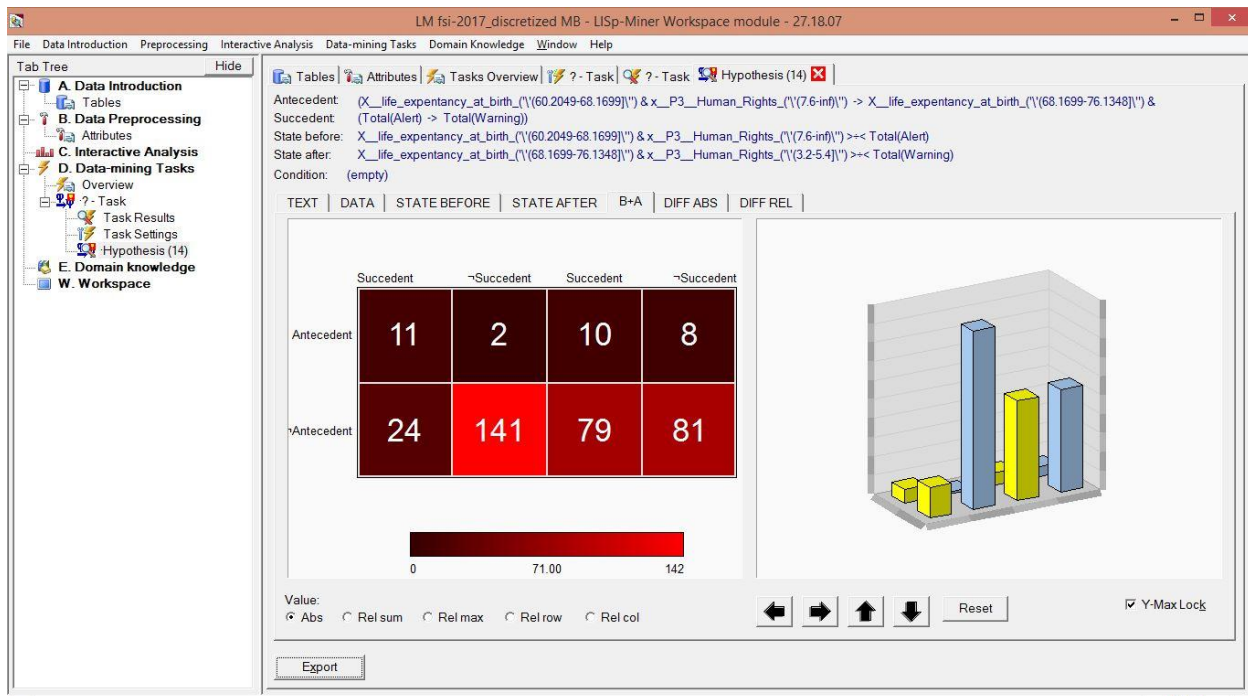
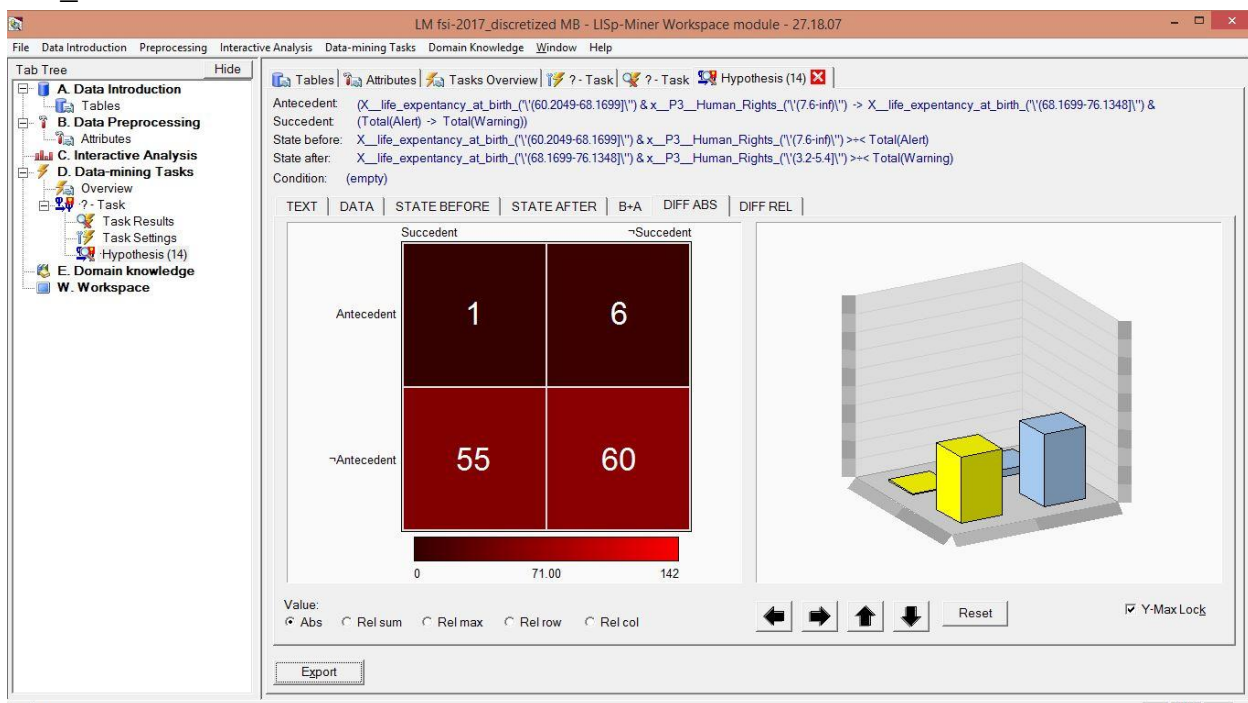
Below is the hypothesis for action rule 1:

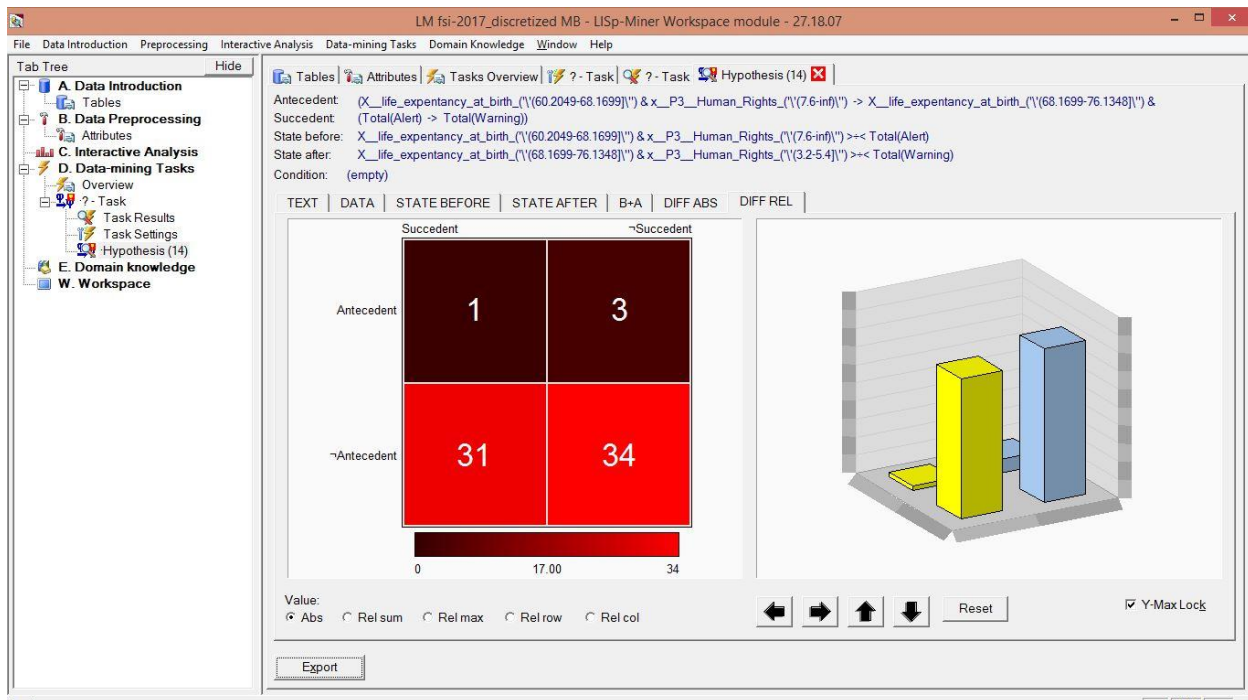
State Before:



State After:



B+A:**Diff_Abs:**

Diff_Rel:**Conclusions**

We added 6 new features in addition to the existing features in the FSI dataset. Those new features were selected that make an impact on the country's FSI value. Then data pre-processing was performed which included data cleaning and normalization. After that, we performed data discretization and data classification using the WEKA tool. Then Lisp Miner was used to generate the action rules.

These action rules will help in improving the condition of a country so that its FSI score improves.

References

- [1] Lisp Miner; <https://lispminer.vse.cz/>
- [2] The Fund for Peace; Fragile States Index; <https://fragilestatesindex.org/>
- [3] Fragile States Index, Wikipedia; https://en.wikipedia.org/wiki/Fragile_States_Index
- [4] Hannah R, Max R; Obesity; University of Oxford; <https://ourworldindata.org/obesity>
- [5] Max R and Mohamed N; Military Spending; University of Oxford; <https://ourworldindata.org/military-spending>

[6] Hannah R, Max R; Hunger and Undernourishment; University of Oxford;
<https://ourworldindata.org/hunger-and-undernourishment>

[7] Our World In Data; Unemployment rate; University of Oxford;
<https://ourworldindata.org/grapher/unemployment-rate>

[8] Data; The World Bank; IBRD-IDA;
<https://data.worldbank.org/indicator/SP.DYN.LE00.IN?end=2018&start=2016>

[9] Women, Peace, and Security Index; Georgetown Institute for Women, Peace, and Security;
<https://giwps.georgetown.edu/the-index/>

[10] Ras; Lisp Miner Manual; Webpages UNCC; <https://webpages.uncc.edu/ras/Paper-AR.pdf>