# 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

## 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.

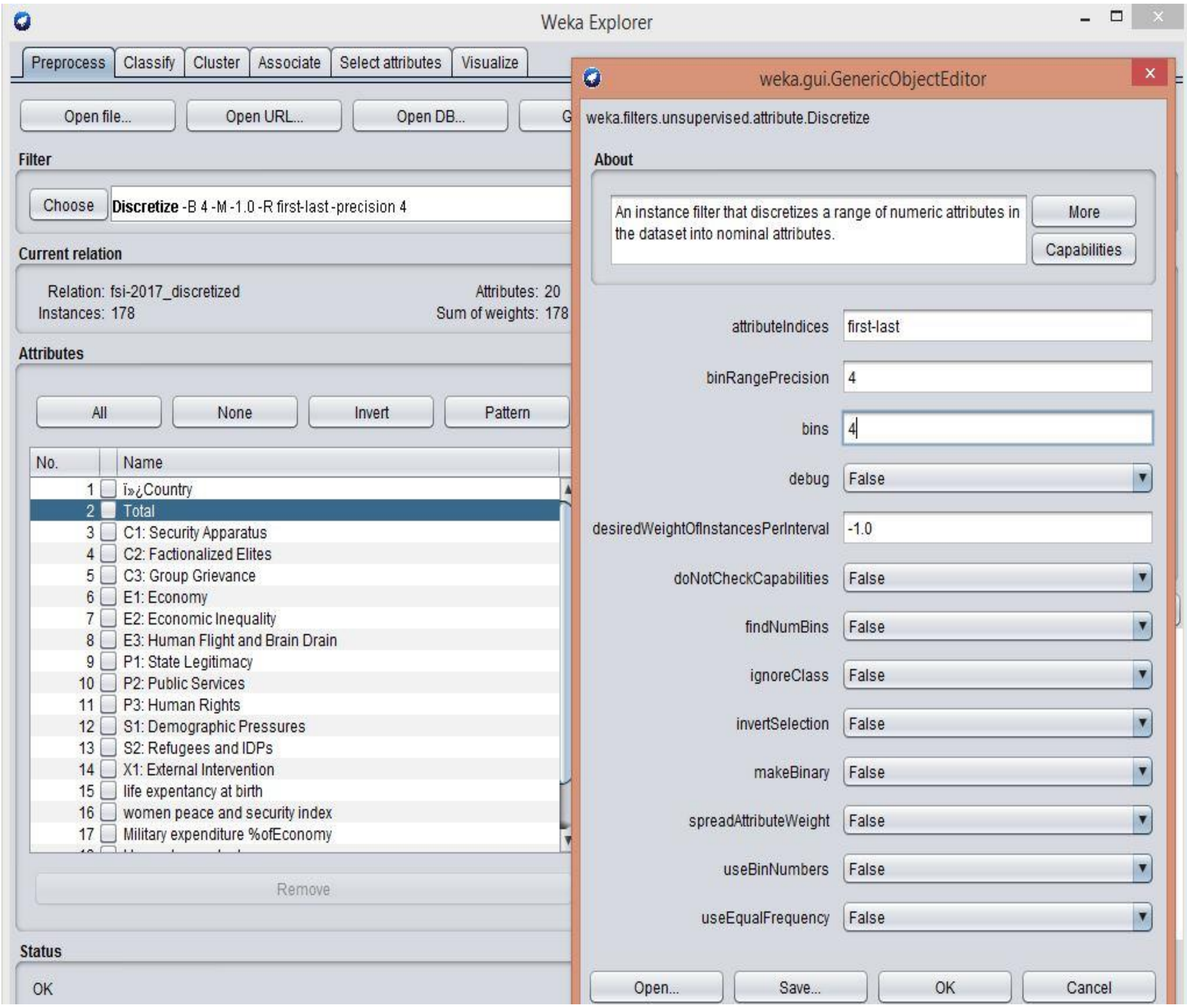
The data analyzed is represented using bit strings that directly results in very fast data mining. The lucid architecture of the system enables the system to be customized for medicine, and finance. The main advantage of the system is to be able to cater to the needs of different professions and the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

## Data Discretization (Weka)

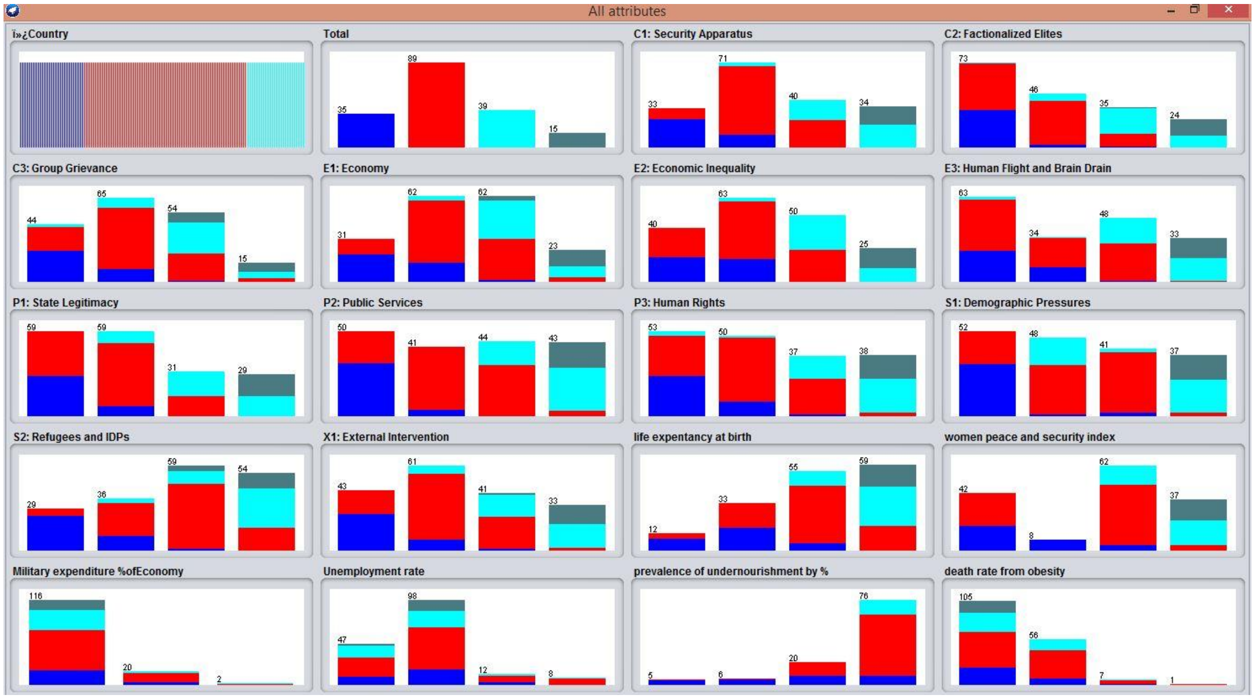
The tool called WEKA is used in this project for the classification of our data.
The discretization is a process that consists of converting or partitioning continuous attributes, features, or variables to discrete or nominal values. The decision attribute TOTAL has continuous values throughout the dataset 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.

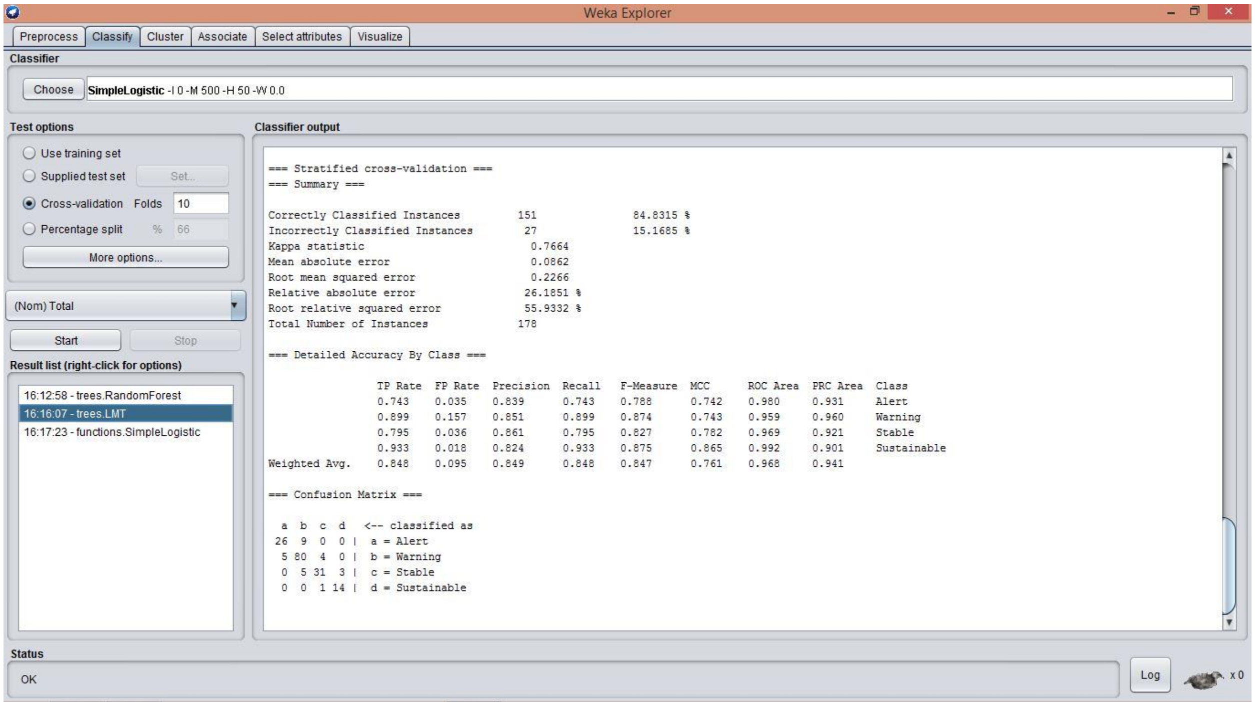
2. **Random Forest:** The Random Forest classification is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual 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.

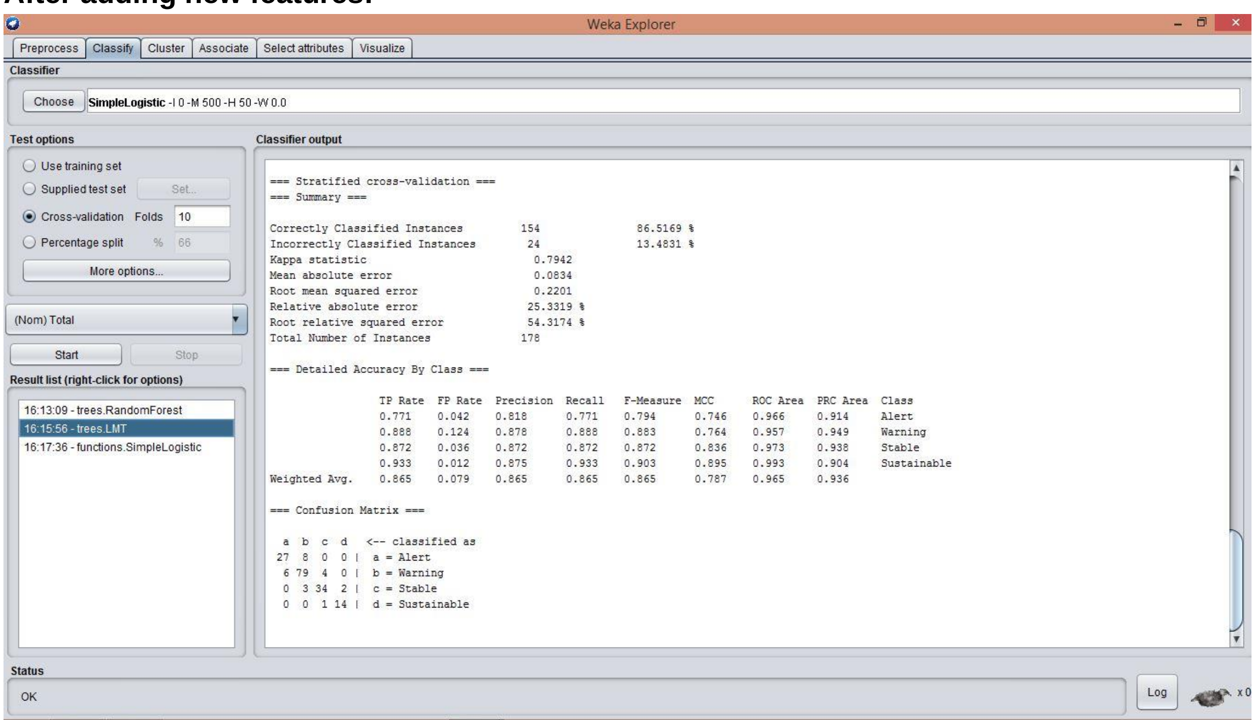
# Weka Screenshots

## 1) Logistic Model Trees

### Before adding new features:

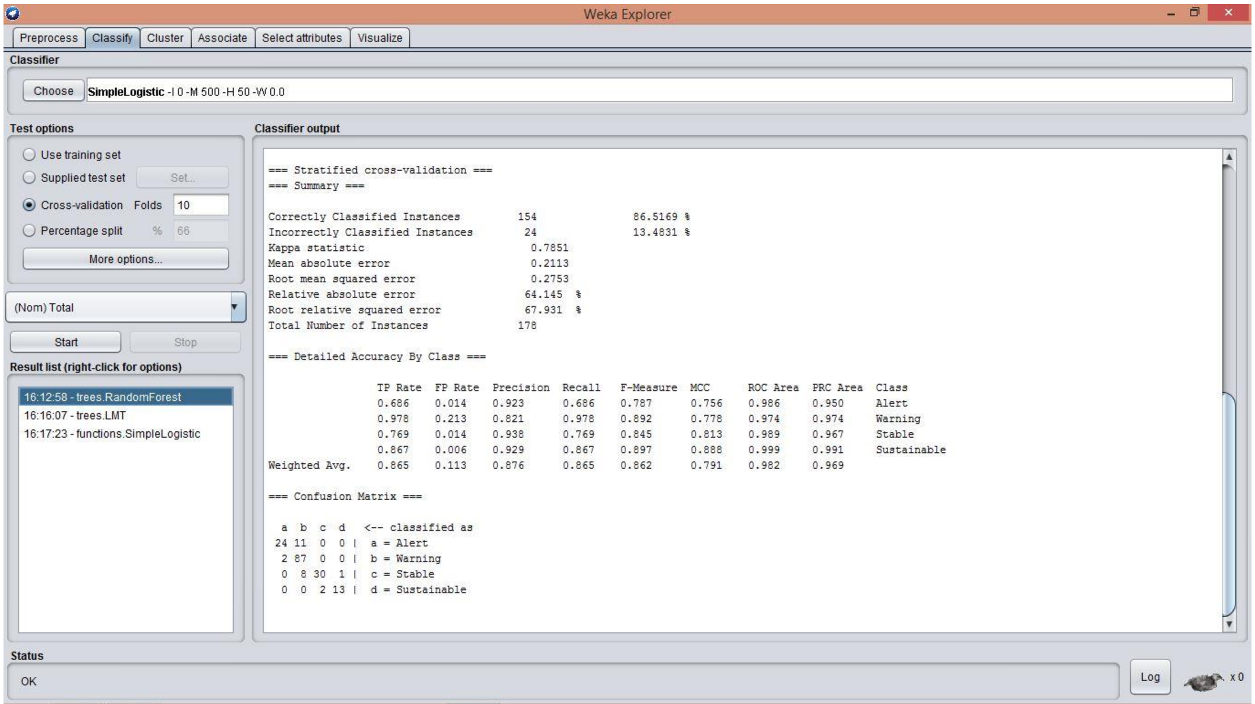



### After adding new features:

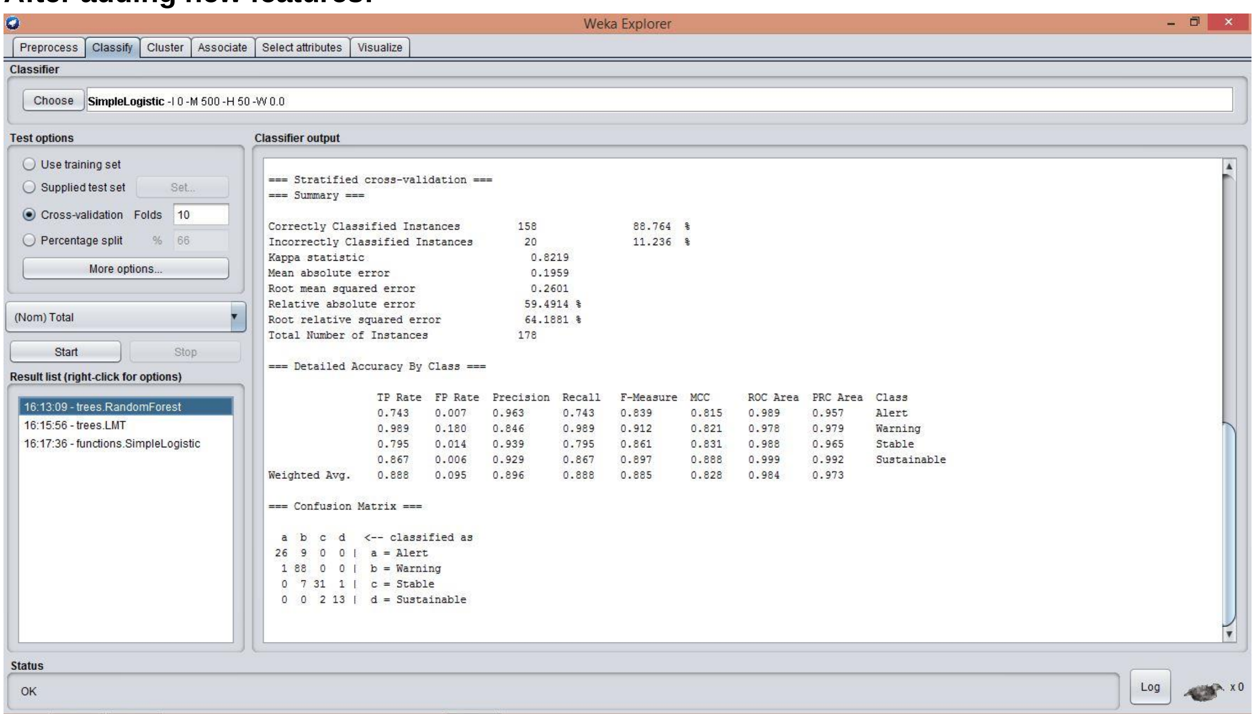


## 2) Random Forest

### Before adding new features:

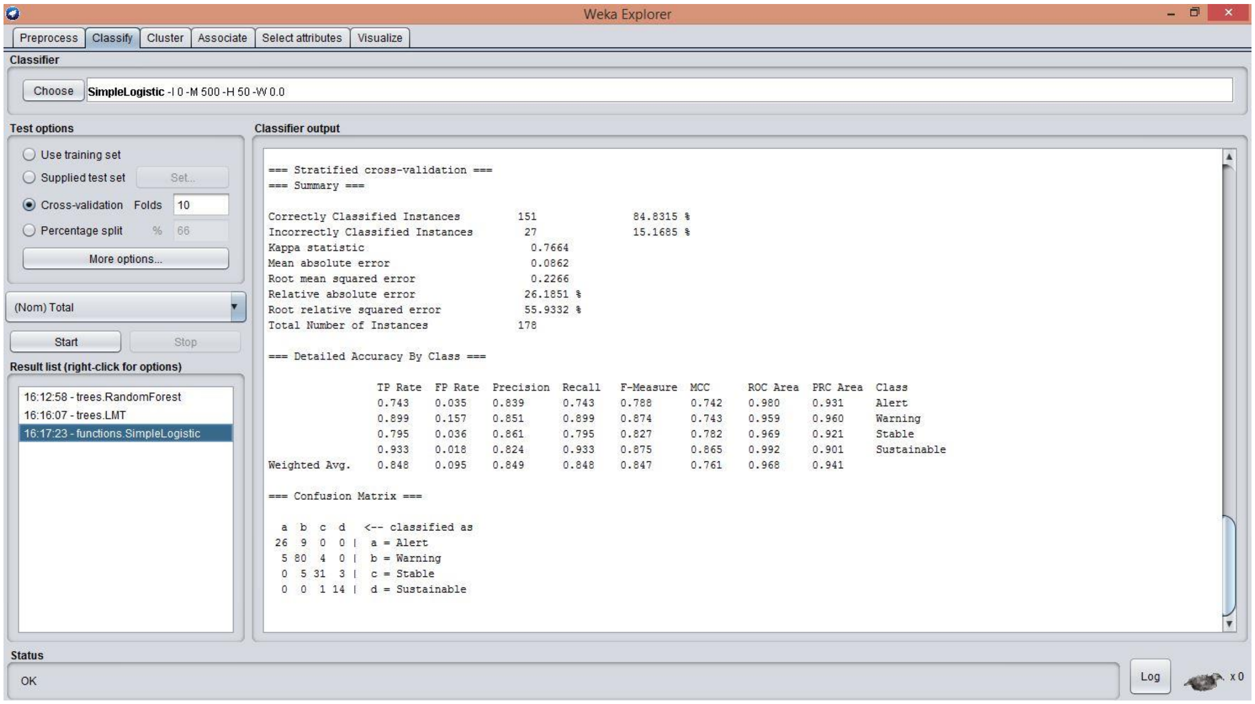



### After adding new features:

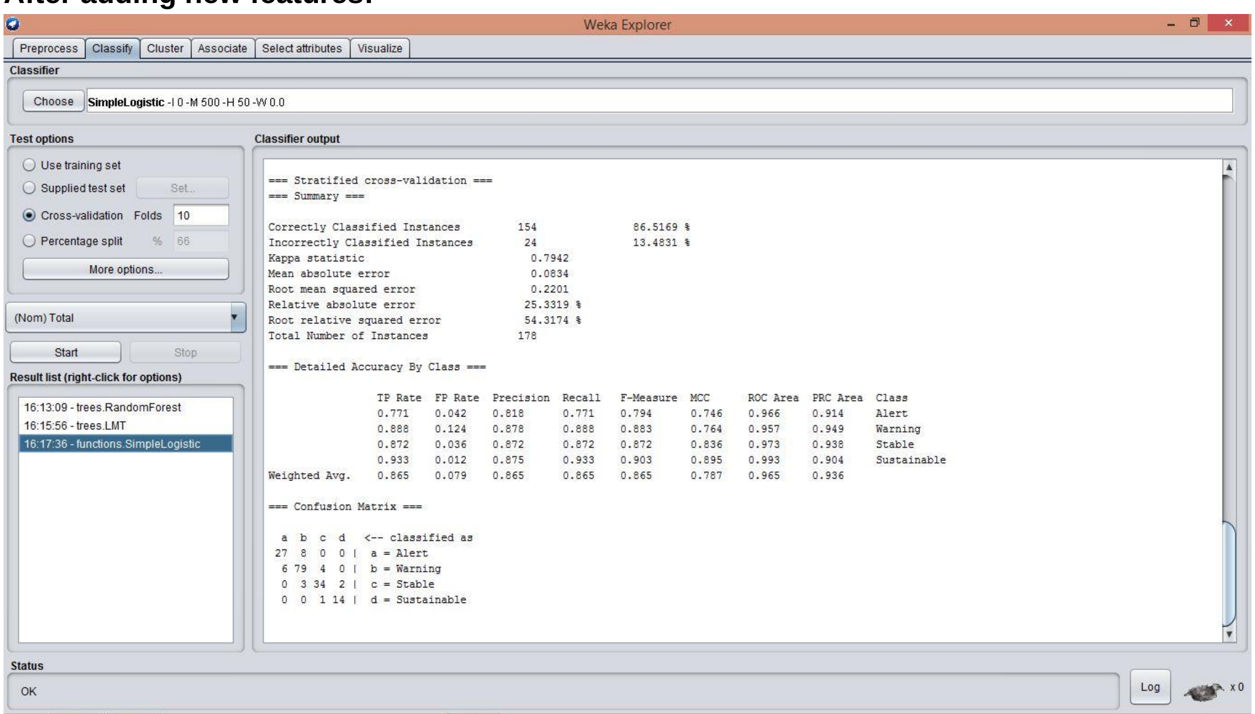


## 3) Simple Logistic

### Before adding new features:




### After adding new features:



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

| ***Classification Algorithms*** | ***Before Adding Attributes*** | ***After Adding Attributes*** |
|---|---|---|
| **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:**

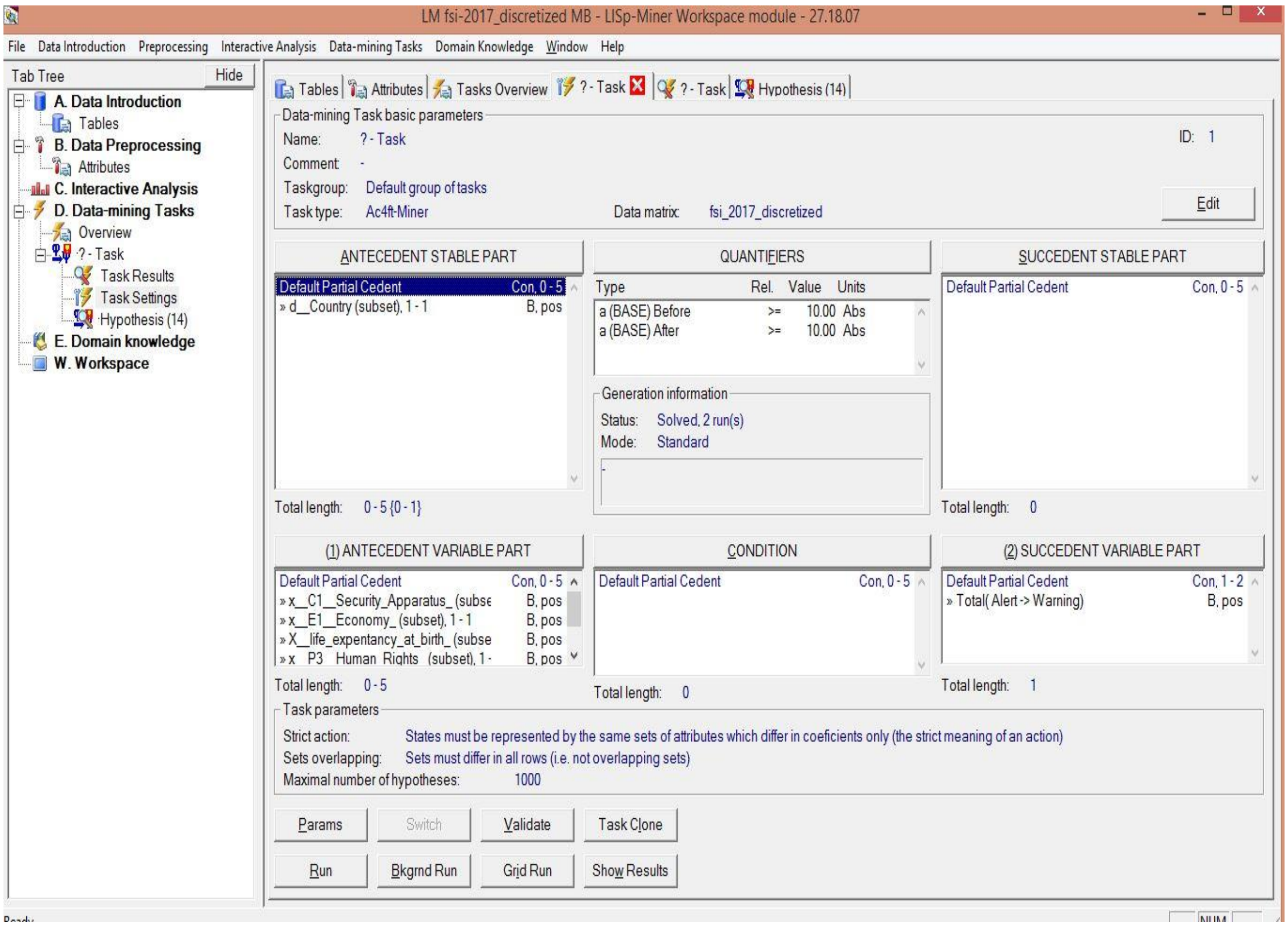
We assigned a decision variable to this set.
Attribute: Total
Coefficient Type: One category

**Quantifiers:**

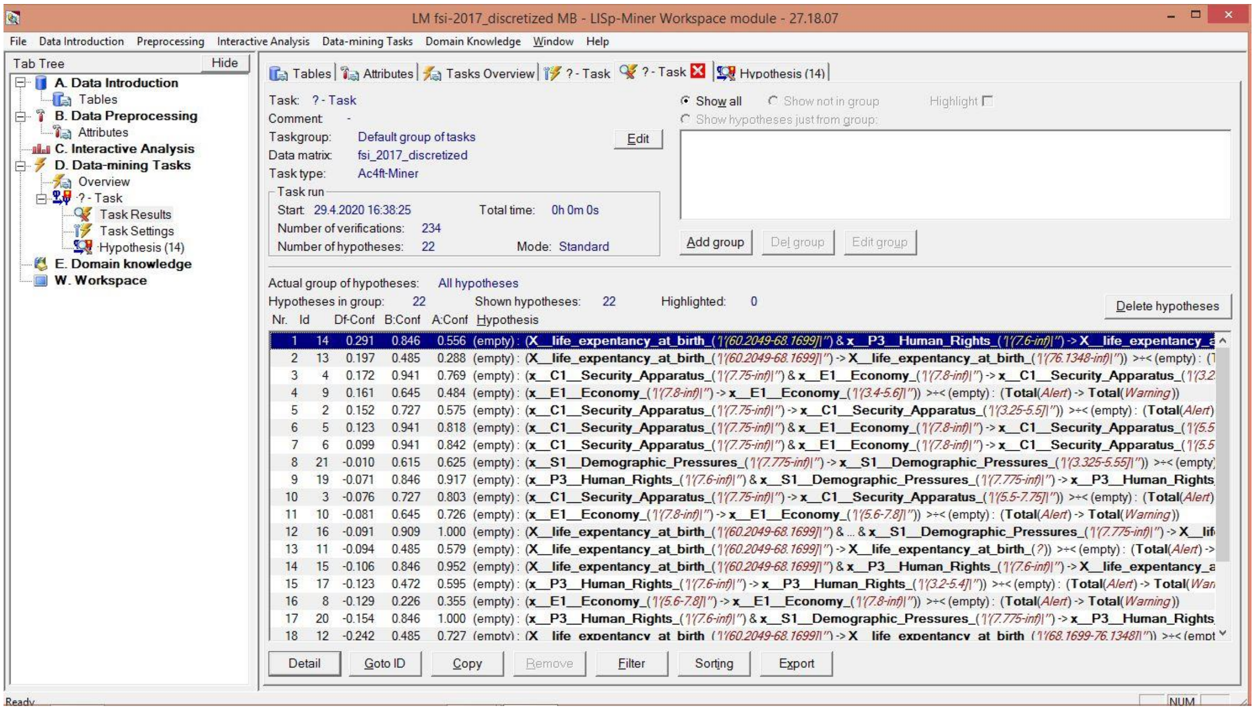
a(Base) Before: 10
a(Base) After: 10

## Extracting Action Rules:

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



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




## 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.

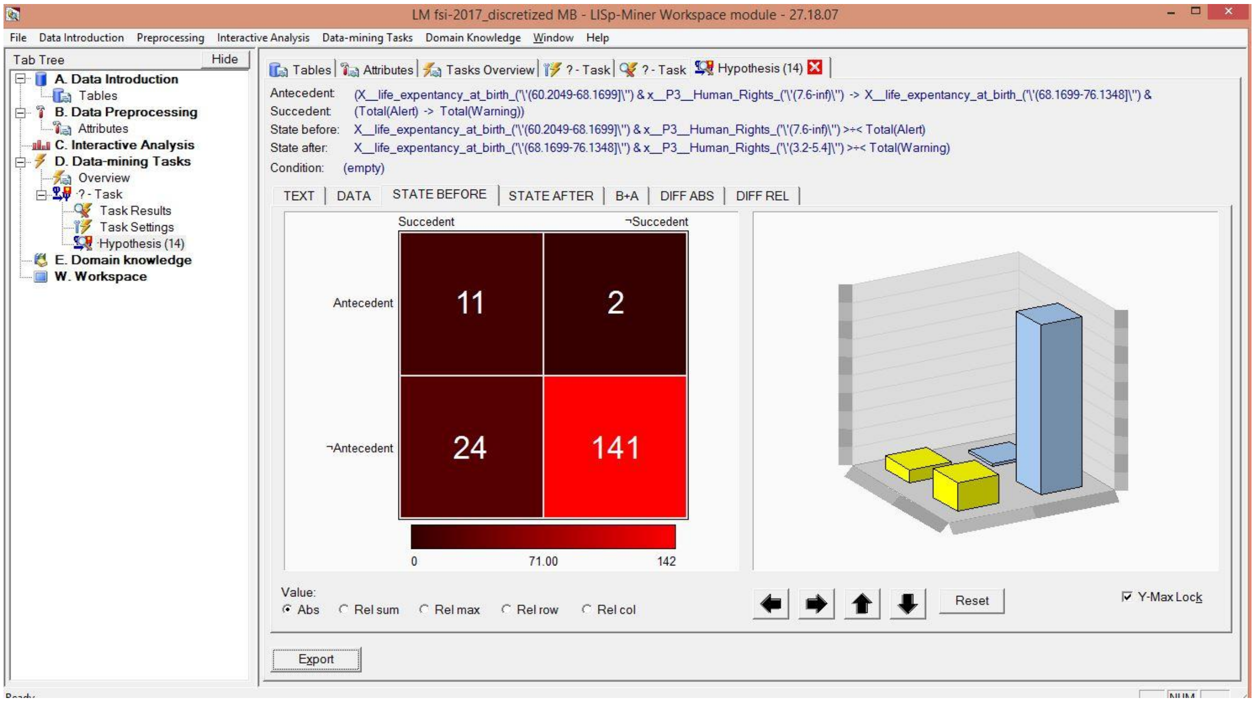
Action Rule 3--- If the Security Apparatus and the Economy of a country is improved according to above-specified metrics then the FSI status of the country can be improved from Alert to Warning.

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.

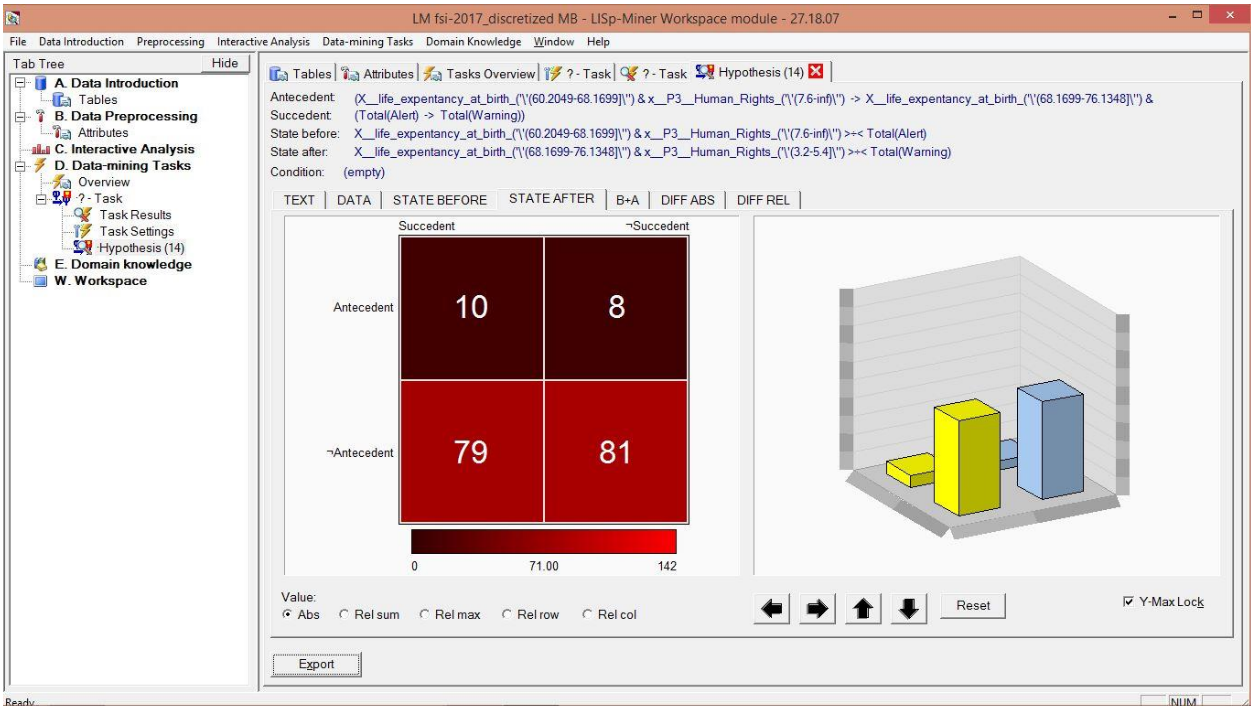
ITCS 6162/8162 Knowledge Discovery in Databases -KDD

Below is the hypothesis for action rule 1:

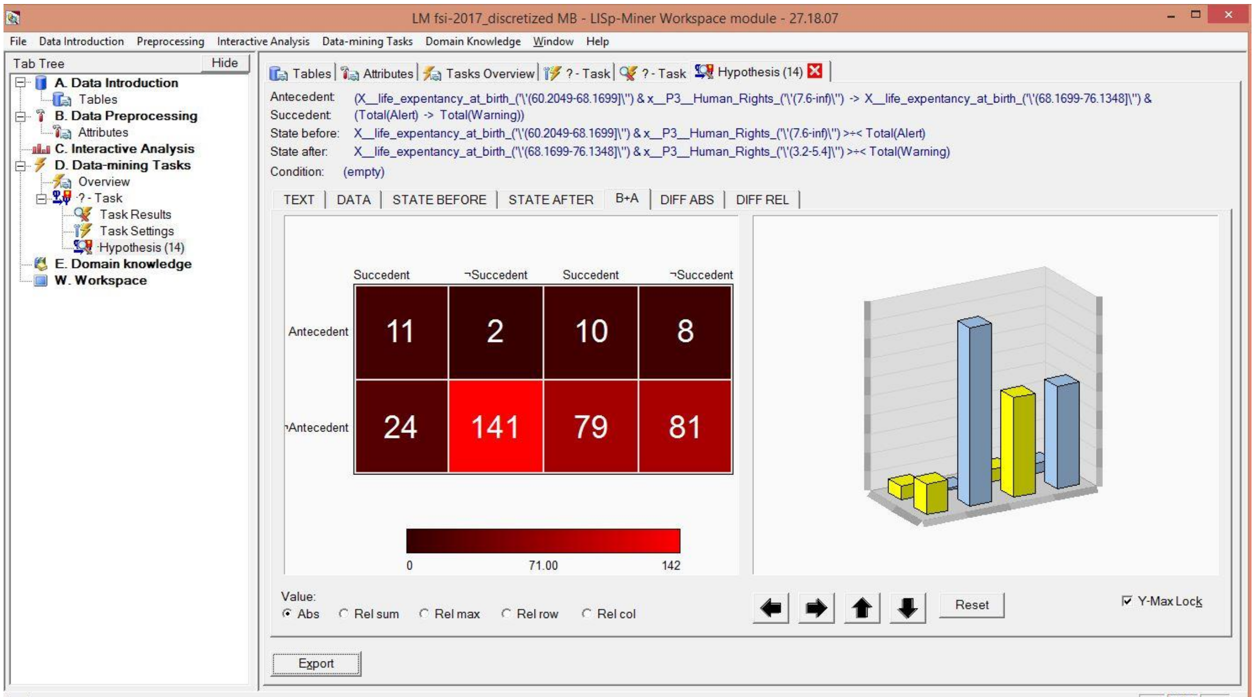
**State Before:**



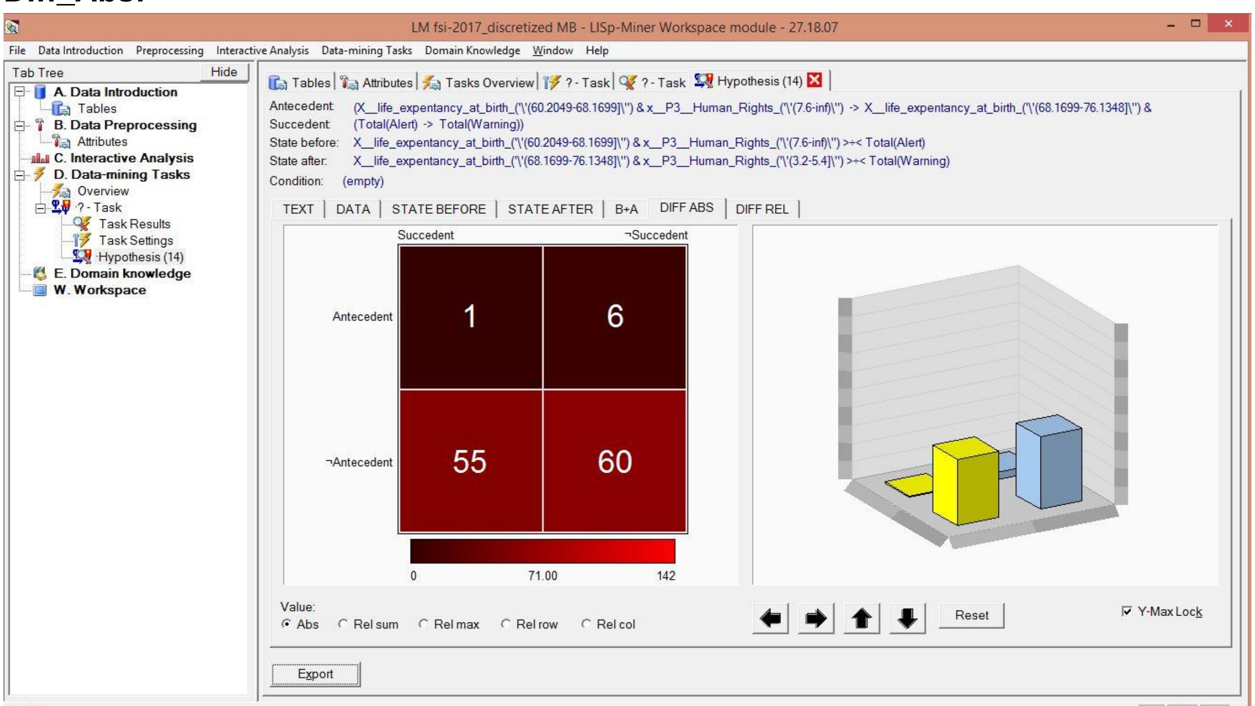

**State After:**



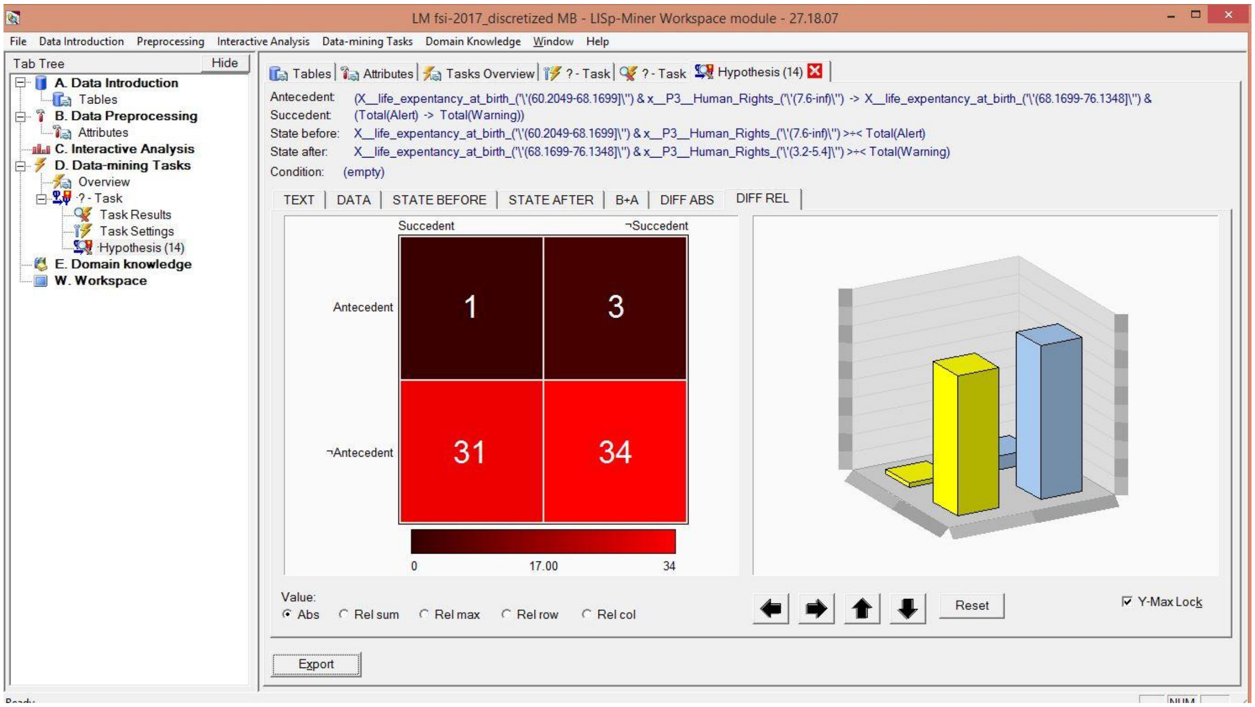
## B+A:


LM fsi-2017_discretized MB - LISp-Miner Workspace module - 27.18.07
File
Data Introduction
Preprocessing
Interactive Analysis
Data-mining Tasks
Domain Knowledge
Window
Help
Tab Tree
Hide
A. Data Introduction
Tables
B. Data Preprocessing
Attributes
C. Interactive Analysis
D. Data-mining Tasks
Overview
? - Task
Task Results
Task Settings
Hypothesis (14)
E. Domain knowledge
W. Workspace
Tables
Attributes
Tasks Overview
? - Task
? - Task
Hypothesis (14)
Antecedent: (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]\'') &
Succedent: (Total(Alert) -> Total(Warning))
State before: X__life_expentancy_at_birth_(\'(60.2049-68.1699]\'') & x__P3__Human_Rights_(\'(7.6-inf)\'') >*< Total(Alert)
State after: X__life_expentancy_at_birth_(\'(68.1699-76.1348]\'') & x__P3__Human_Rights_(\'(3.2-5.4]\'') >*< Total(Warning)
Condition: (empty)
TEXT
DATA
STATE BEFORE
STATE AFTER
B+A
DIFF ABS
DIFF REL
Succedent
¬Succedent
Succedent
¬Succedent
Antecedent
¬Antecedent
11
2
10
8
24
141
79
81
0
71.00
142
Value:
Abs
Rel sum
Rel max
Rel row
Rel col
Reset
Y-Max Lock
Export

## Diff_Abs:


LM fsi-2017_discretized MB - LISp-Miner Workspace module - 27.18.07
File
Data Introduction
Preprocessing
Interactive Analysis
Data-mining Tasks
Domain Knowledge
Window
Help
Tab Tree
Hide
A. Data Introduction
Tables
B. Data Preprocessing
Attributes
C. Interactive Analysis
D. Data-mining Tasks
Overview
? - Task
Task Results
Task Settings
Hypothesis (14)
E. Domain knowledge
W. Workspace
Tables
Attributes
Tasks Overview
? - Task
? - Task
Hypothesis (14)
Antecedent: (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]\'') &
Succedent: (Total(Alert) -> Total(Warning))
State before: X__life_expentancy_at_birth_(\'(60.2049-68.1699]\'') & x__P3__Human_Rights_(\'(7.6-inf)\'') >*< Total(Alert)
State after: X__life_expentancy_at_birth_(\'(68.1699-76.1348]\'') & x__P3__Human_Rights_(\'(3.2-5.4]\'') >*< Total(Warning)
Condition: (empty)
TEXT
DATA
STATE BEFORE
STATE AFTER
B+A
DIFF ABS
DIFF REL
Succedent
¬Succedent
Antecedent
¬Antecedent
1
6
55
60
0
71.00
142
Value:
Abs
Rel sum
Rel max
Rel row
Rel col
Reset
Y-Max Lock
Export

**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