**Analysis on Fragile States Data**



**REPORT**

FINAL PROJECT

**ITCS 6162 - Knowledge Disc in Databases**

SUBMITTED TO

Dr. [Zbigniew W. Ras](https://webpages.uncc.edu/ras/" \t "https://uncc.instructure.com/courses/_blank), Ph.D.

SUBMITTED BY:

|  |  |
| --- | --- |
| RISHAB SEMLANI | 801201960 |
| VARUN JAIN | 801168344 |
| DHRUVIL PATEL | 801219704 |
| DARSHAN KRISHNAPURA KRISHNA | 801217746 |



**Table of Contents**

1. **Project Description and Requirements 3**
   1. Fragile State Index 4
   2. Indicators 4
2. **Extended Features 4**
3. **Decision Attributes 15**
4. **Classification Using Weka 15**
   1. Original dataset 4
   2. Extended (additional Features) dataset 4
5. **Generation of action rules using Lisp Miner 15**
6. **Lisp Miner Screenshots 15**
7. **Fragile Attributes Used Year Wise 15**
8. **Action rules and Inferences 15**
9. **Data sources and methods of extractions 15**

**10 Conclusion 15**

**11 References 15**

**Project Description and Requirements:**

The main goal of this project is to conduct action rule mining for the given Fragile State Index (FSI) data sets with new extended features (six features) added to the original data set, as well as further research to determine what improvements to the classification features are needed to lower the FSI.

Using Action-Rules, we decide how a nation can transition from an alert to a stable state.

**Fragile State Index**

The Fragile States Index (FSI) is an annual report published since 2005 by the United States think tank of the Fund for Peace and the American Foreign Policy magazine. The list is intended to assess the vulnerability of states to conflict or collapse, ranking all sovereign states with United Nations membership where sufficient data are available for analysis. In spite of being recognized as sovereign by one or more other countries, Taiwan, the Palestinian Territories, Northern Cyprus, Kosovo and Western Sahara are not listed. Ranking is based on the 12 criteria number of ratings. Each indicator is scored on a scale of 0 to 10, with 0 being the lowest (most stable) intensity and 10 being the highest (least stable) intensity, creating a scale of 0−120.

The fragile state of each country depends on its score.

Alert – 90.0 to 120.0

Warning – 60.0 to 89.9

Stable – 30.0 to 59.9

Sustainable – 0.0 to 29.9

**Indicators:**

At any given moment, twelve conflict risk metrics are used to assess a state's situation. The metrics provide a timely snapshot that can be calculated in a time series against other snapshots to assess whether conditions are improving or worsening. The list of indicators used in both the CAST framework and the Fragile States Index is as follows:

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

**Extended Features:**

1. **Unemployment Rate:**

Unemployment refers to the job-power offer, which is however accessible to and seeking employment without work. The higher the unemployment in the country, the more problems are caused. The unemployed youth may engage in unlawful practices to purchase money, thereby increasing the nation's fragility.

1. **Total Tax Rate:**

In the after exemptions and allowable deductions are made, the total tax rate estimates the measure of taxes and required contributions payable by organizations. Taxes withheld, such as personal taxes and taxes, such as sales / goods tax, are excluded. Higher tax rates mean low fragility for angry customers.

1. **Military Expenditure:**

Based on NATO concepts, this SIPRI information is derived and standardized. This data includes all spending and capital currently invested in the defense, including peacekeeping forces, defense ministries, and other government agencies for defense purposes. Peaceful nations like Sweden wouldn't have to spend much on the military because there's not a lot of wars.

1. **Inflation:**

Inflation is a metric that is a statistical trend which raises the average cost over a span of a collection of goods and services. It is a slow and steady increase in price rates where a unit currency can purchase fewer goods / services than in previous years. Higher inflation can lead to higher fragility in the public outcry.

1. **Population growth:**

Population growth refers to a rise in a country's number of people. This data applies to the amount of the rise from the previous year in the current year. Higher population growth leads to higher inflation, military spending, unemployment rate, etc., resulting in increased fragility.

1. **Access to electricity:**

This amount represents the percentage of the population with electricity exposure. A ranking of 100 percent means a higher living standard. Therefore, to get a score on 1-10, this percentage is subtracted from 100 and divided by 10. Countries with 100% electricity access would thus have a null rating.

**Decision Attributes**

As described in the wiki page about the fragile index, using 12 indicators the total score is then used to classify countries.

The ranges defined 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 6 new attributes the score was normalized in the following way:

* The total was reduced from 180 (18 indicators each with a score from 1-10) to 120.
* Example a score of 60 on 180 would imply 60\*120/180 = 40 on 120

The 6 new indicators introduced, however, did not resonate with the existing 12 indicators, i.e. the number was slightly lower than expected after adding 6 new indicators.

The ranges were therefore redefined to accommodate the disparities, this range was determined by analyzing the original datasets and attempting to group similar counts of countries in the fragile states.

The new ranges were

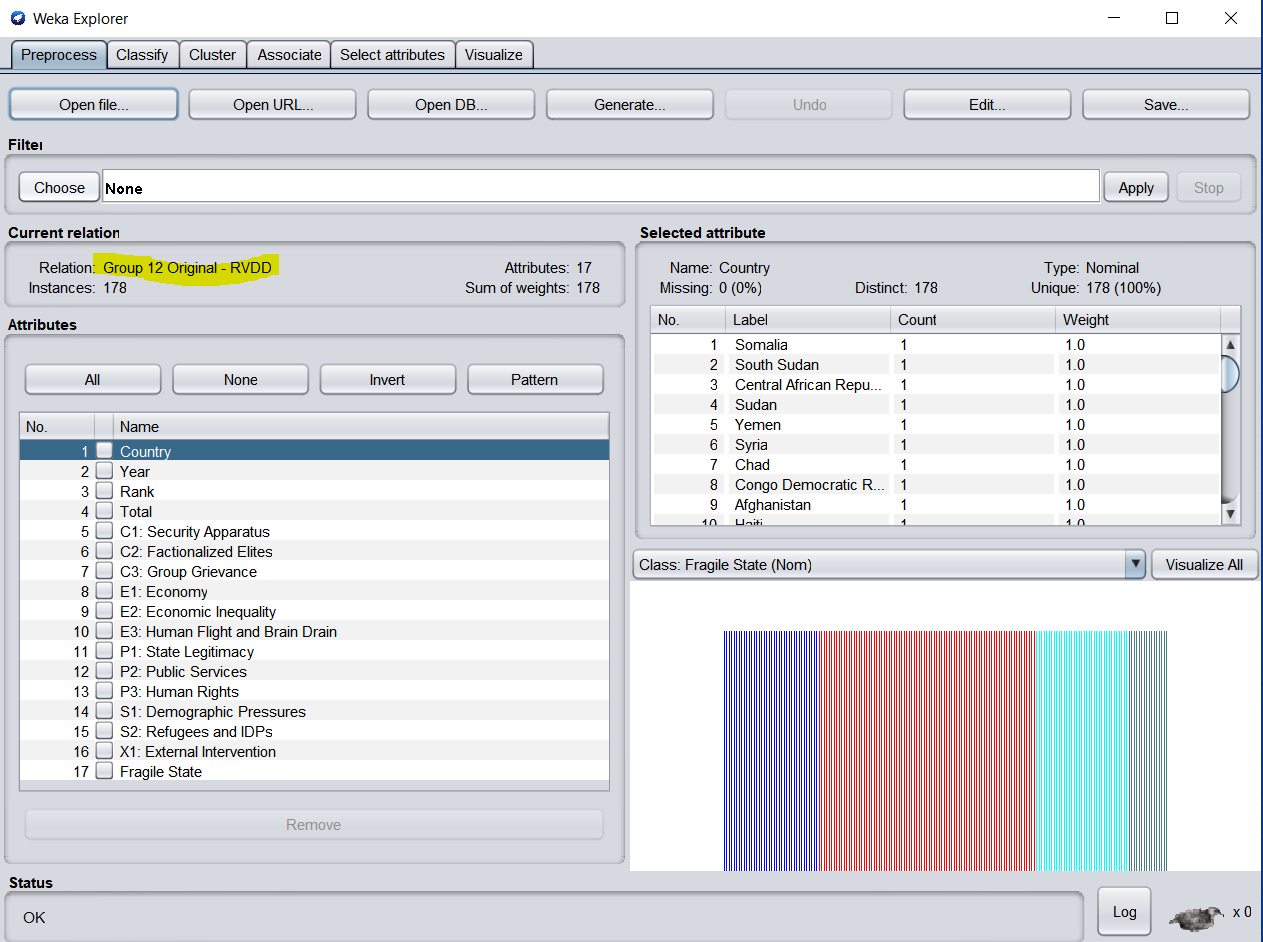
* Alert (>70.0)
* Warning (45.0 – 69.9)
* Stable (25.0-44.9)
* Sustainable (0-25)

This classification was used for the extended data sets.

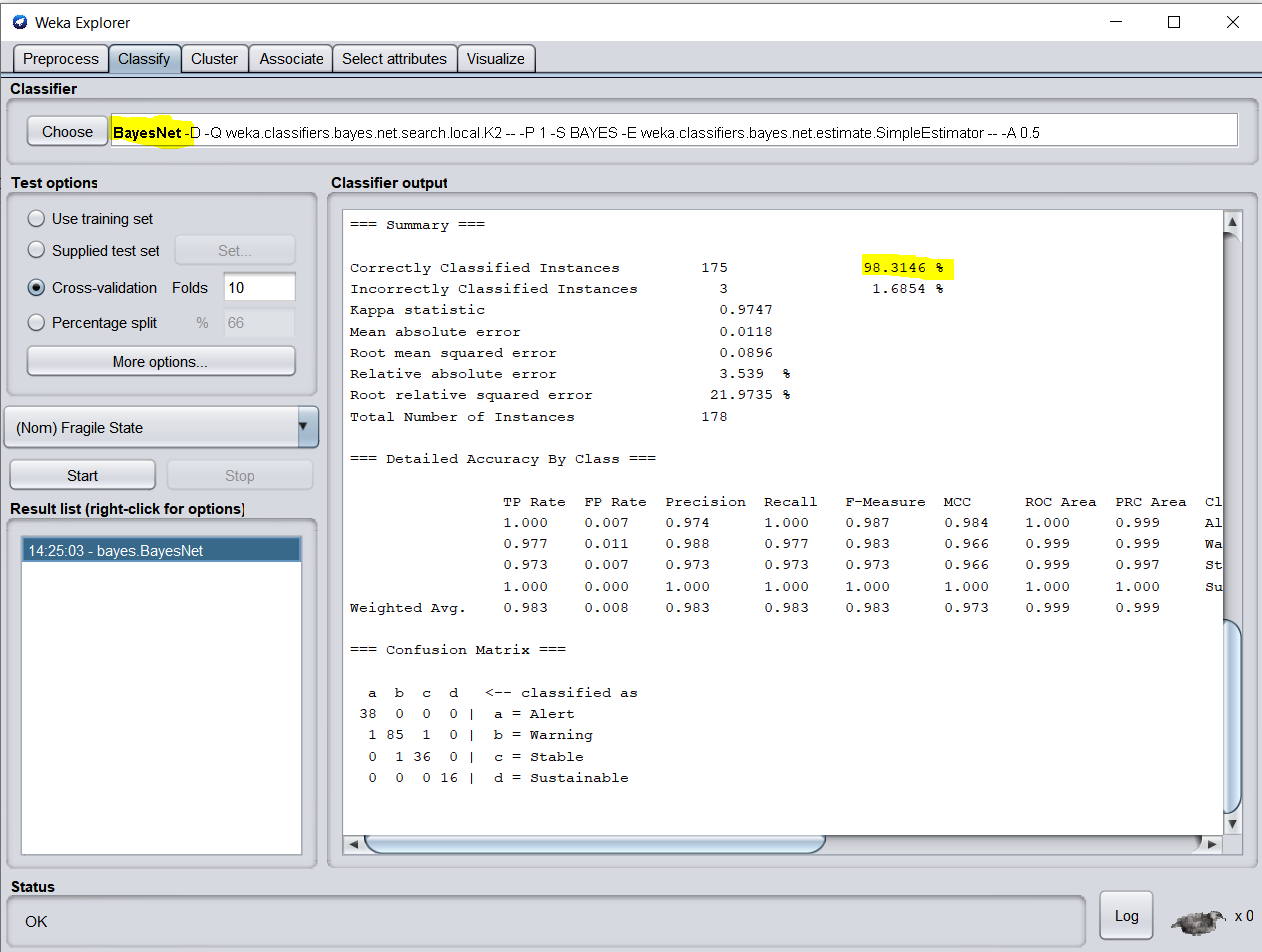
**Classifiers Using WEKA:**

We took 2016 data to perform the following experiment as all the extended data were readily available.

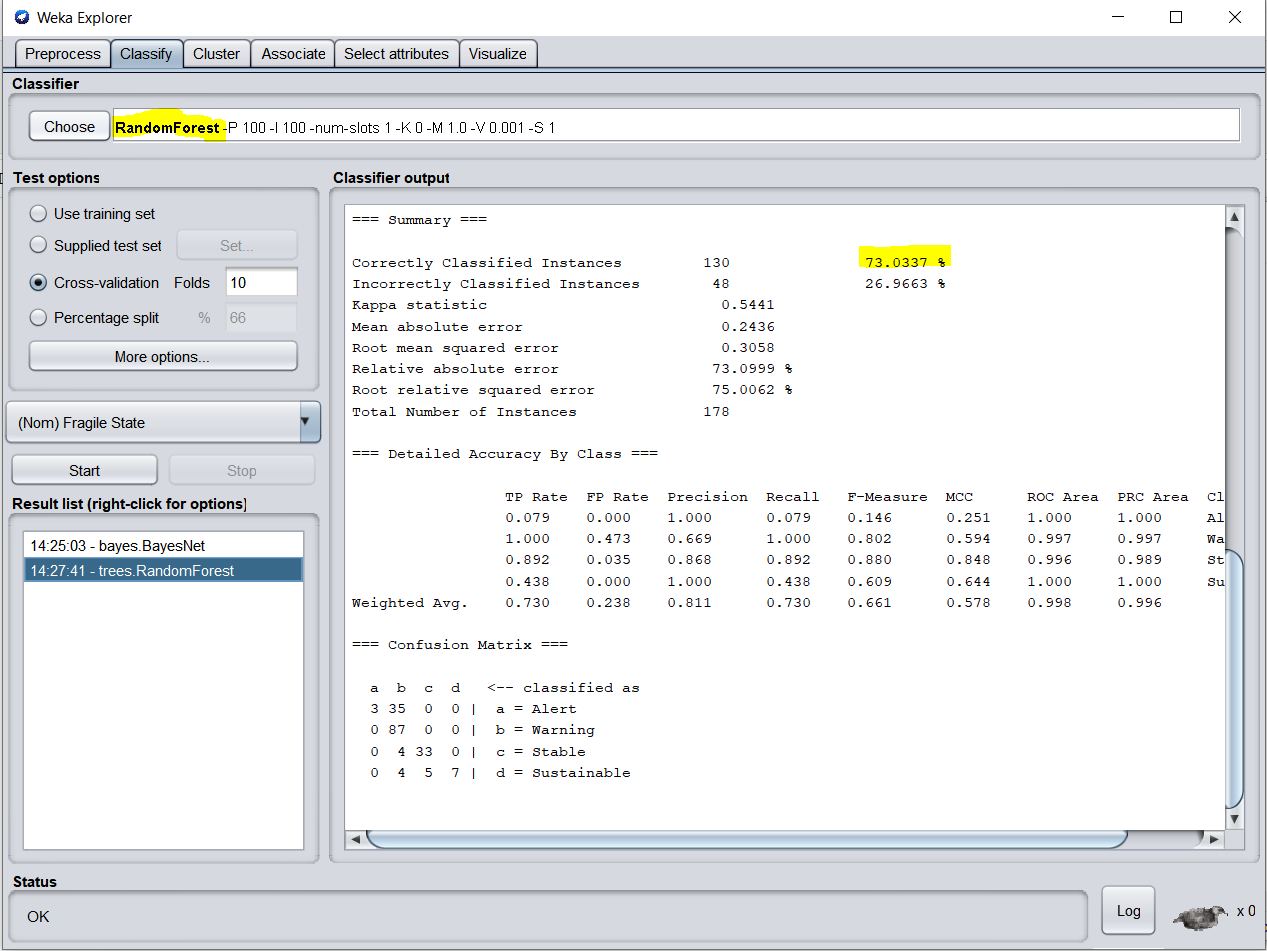
**Original Dataset:**



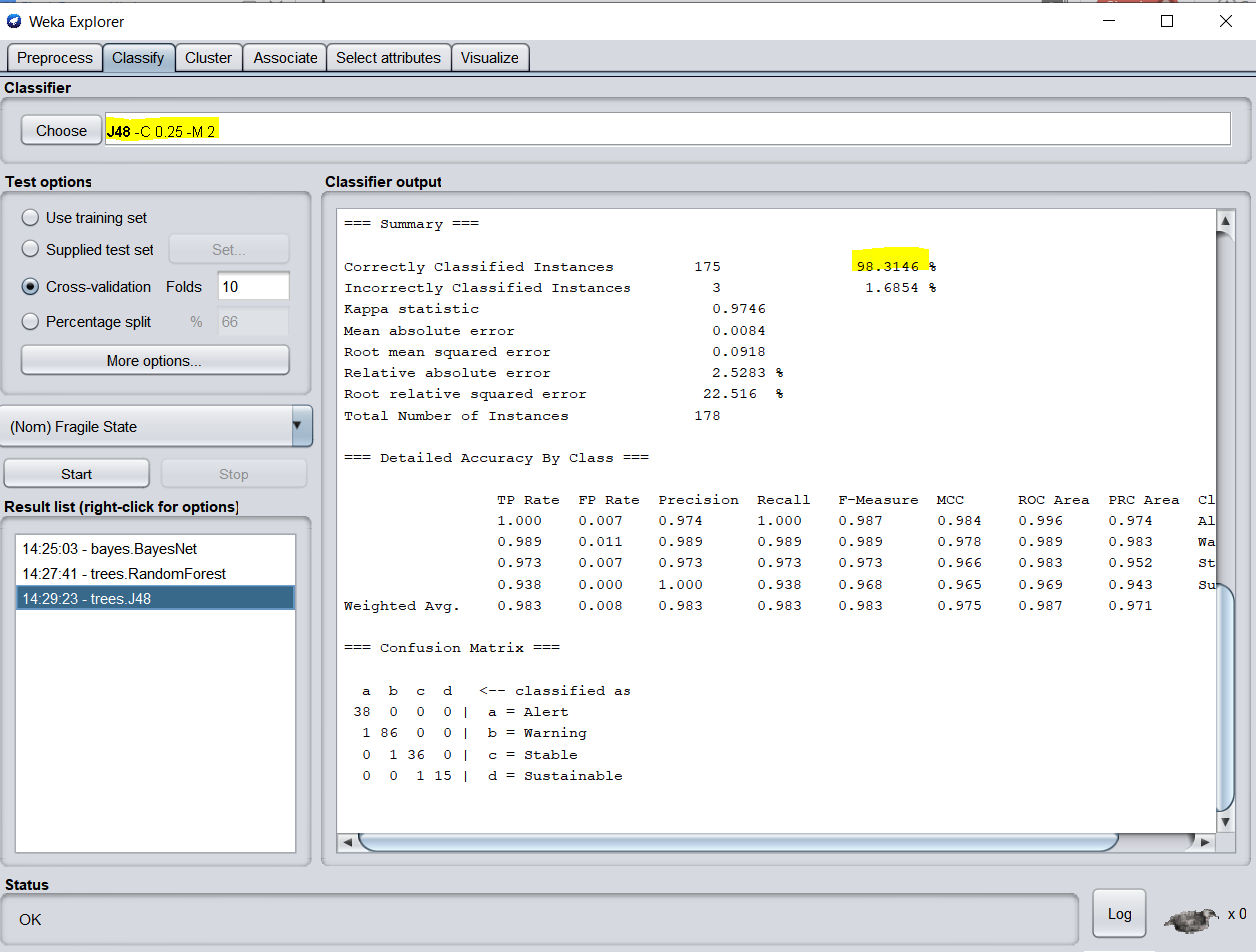
Original data of 2016 imported in WEKA to perform further analysis.



BayesNet Classification on Original Data Set

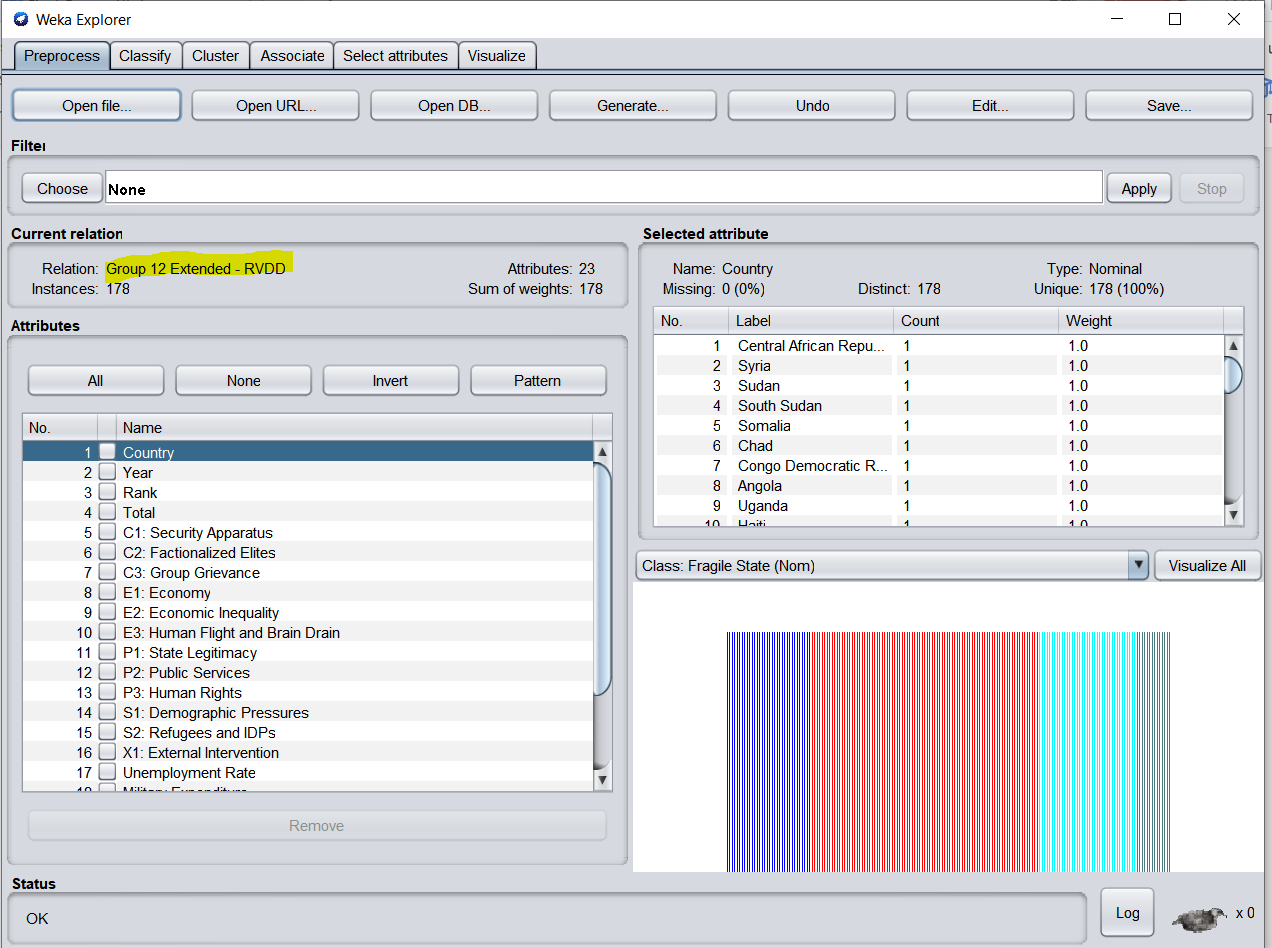


RandomForest Classification on Original Data Set

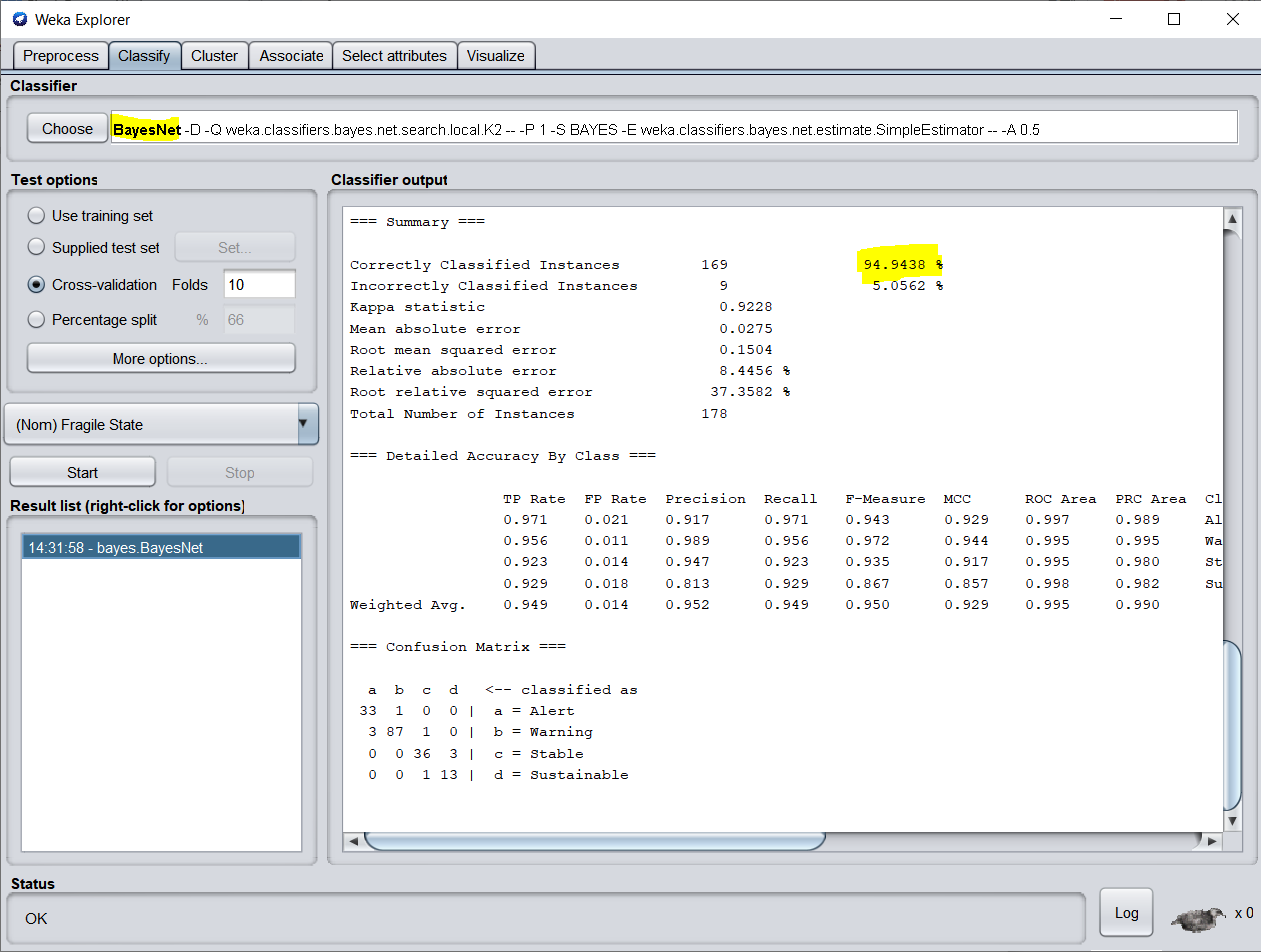


J48 Classification on Original Data Set

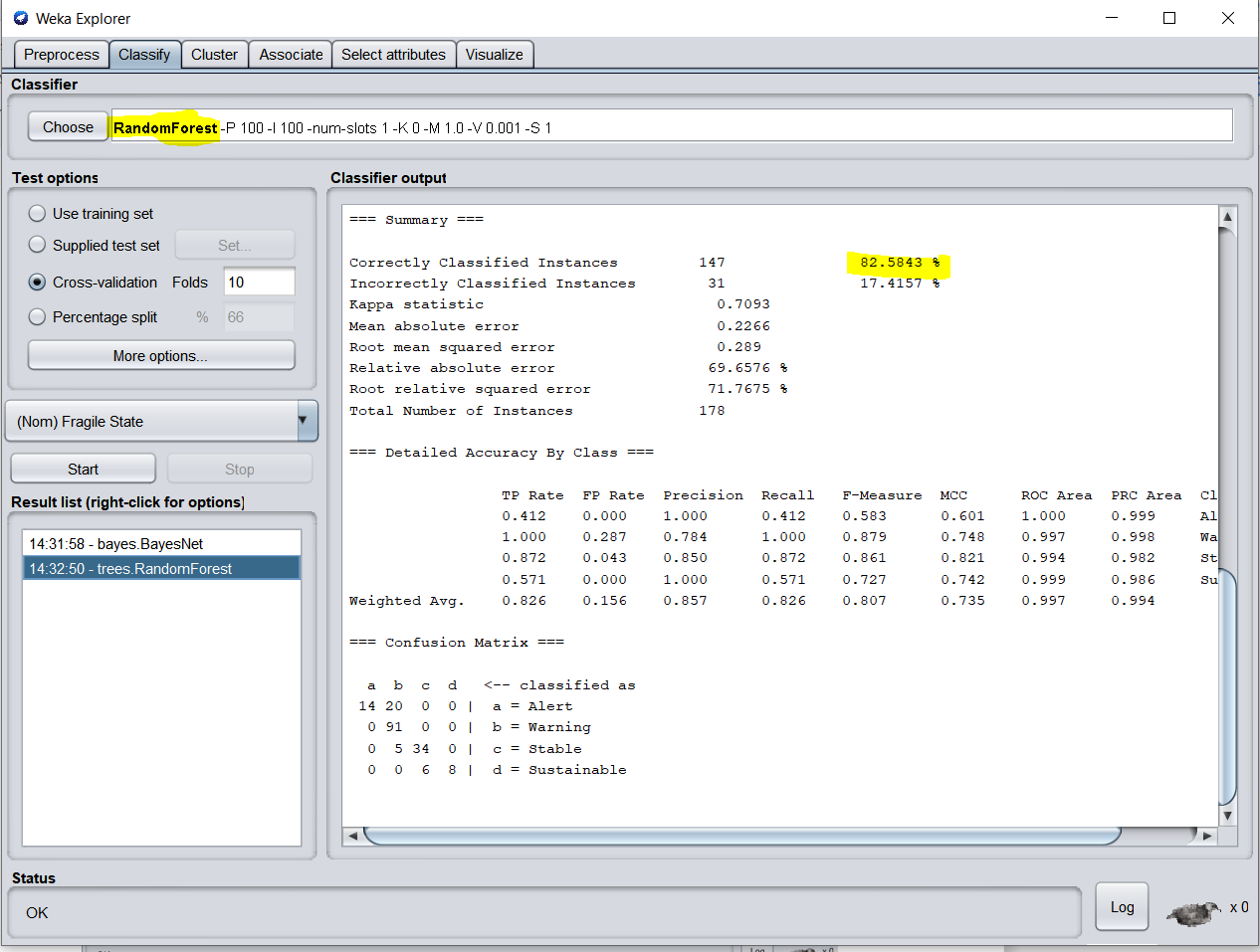
**Extended Data Set:**



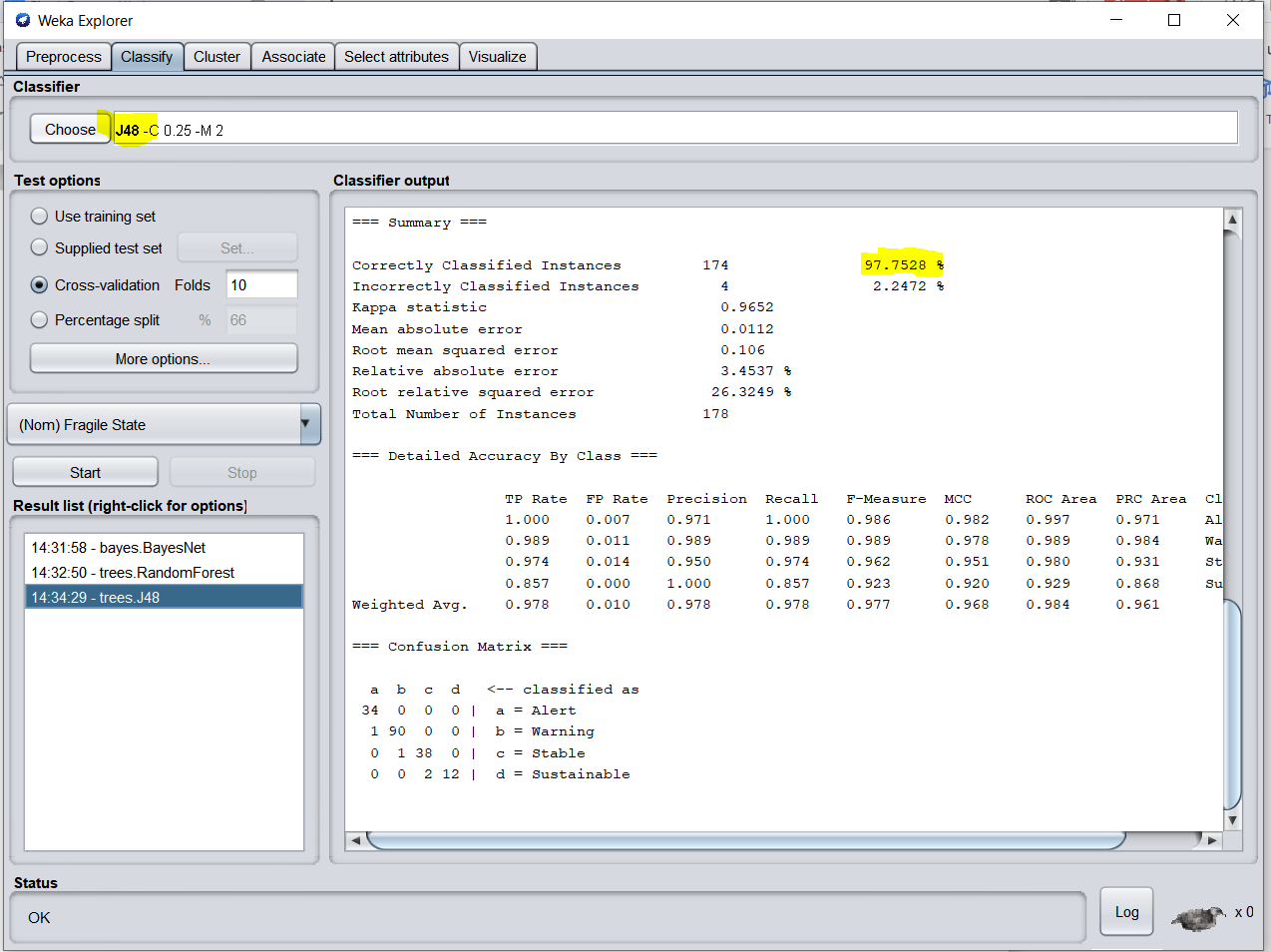
Extended data of 2016 imported in WEKA to perform further analysis.



BayesNet Classification on Extended Data Set



RandomForest Classification on Extended Data Set



J48 Classification on Extended Data Set

For the year 2016, the following table was built on the basis of the results obtained by running the classifiers on original data set.

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Bayes Net | Random Forest | J48 |
| 2016 | 98.31 | 73.03 | 98.31 |

Based on the results obtained from running the classifiers on extended data sets for the year 2016 the following table was built:

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Bayes Net | Random Forest | J48 |
| 2016 | 94.94 | 82.58 | 97.75 |

**Conclusion:**

From the above results we can infer that with higher consistency, the J48 classifier is better. We can note that for the initial datasets, it offers 98.31%. We also have the result for the expanded data sets as 97.75%. To provide predictable results for both initial and extended data **best Classifier Algorithm for the dataset is J48.**

**Generation of Action Rules using Lisp Miner:**

**Properties** :

We have grouped our qualities into 3 classes –

* Stable Attributes
* Flexible Attributes
* Decision Attributes

We have picked the steady properties as :

* Antecedent stable part :

We dole out all the steady ascribes to this set.

* Antecedent variable part :

We allocate all the adaptable ascribes to this set.

* Succedent variable part :

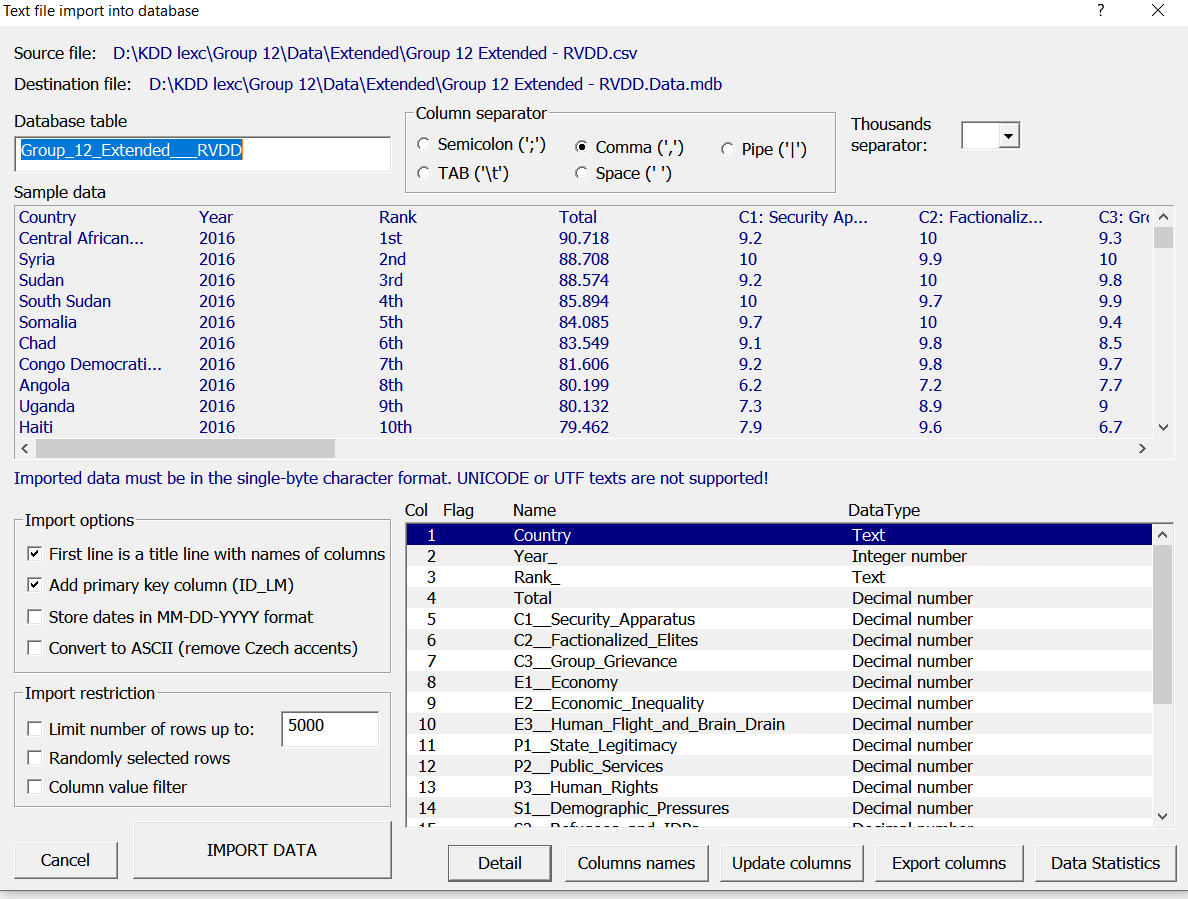
We assign the decision variable to this set.

* Attribute type = nominal
* Coefficient type = one category

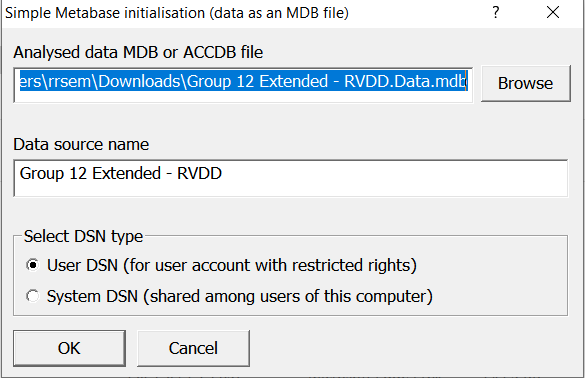
**Quantifiers** :

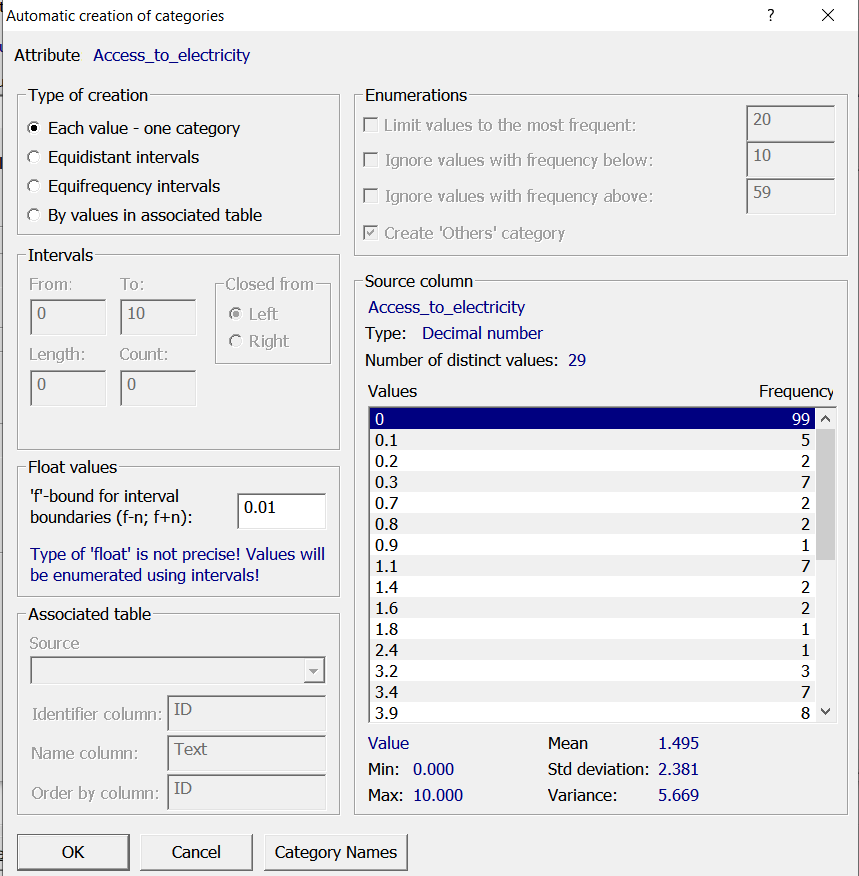
* a(BASE)Before : 2
* a(BASE)After : 2

**Lisp Miner Screenshots:**

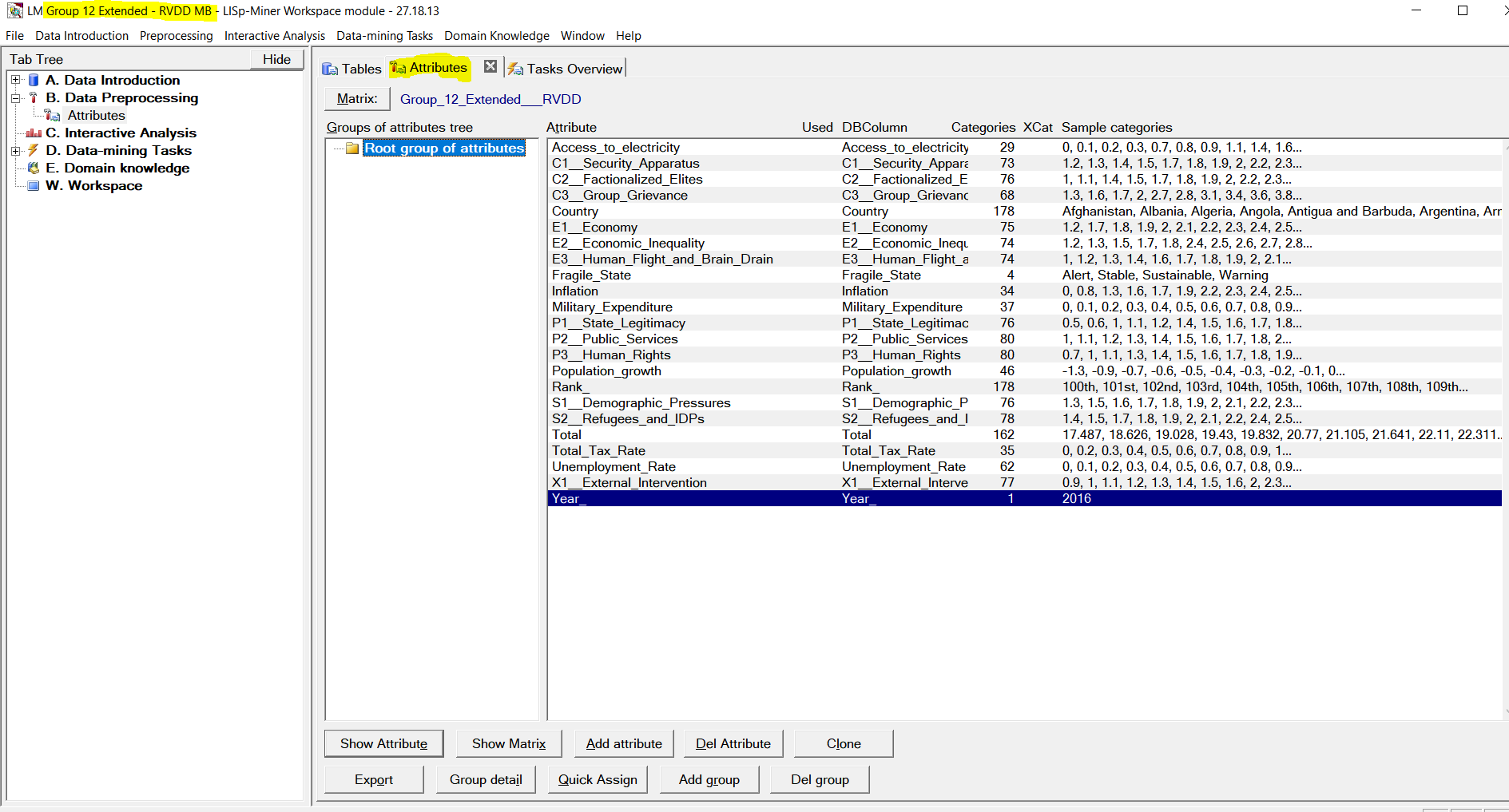


Extended data of 2016 imported in LISP Miner to perform further analysis.

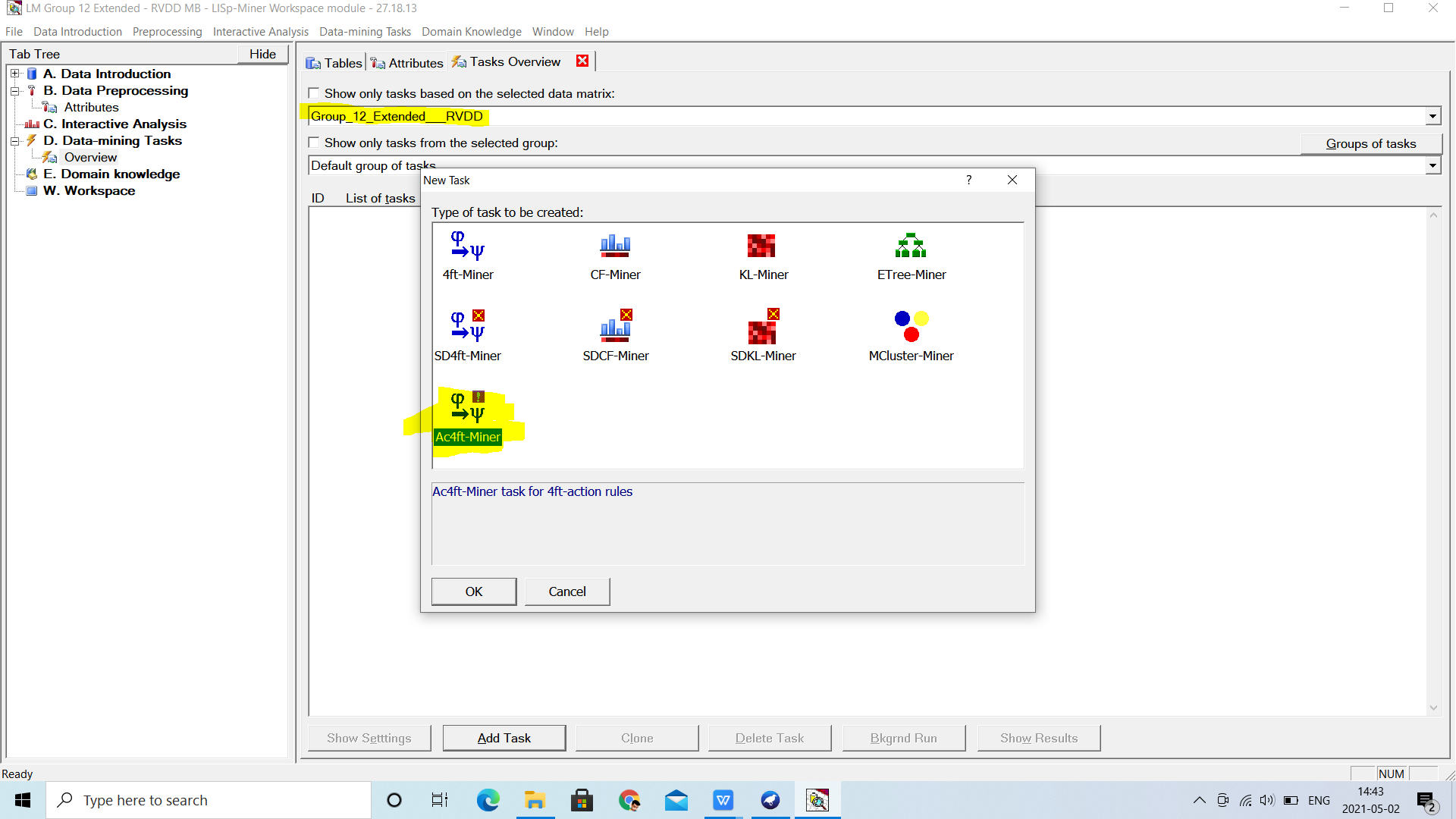




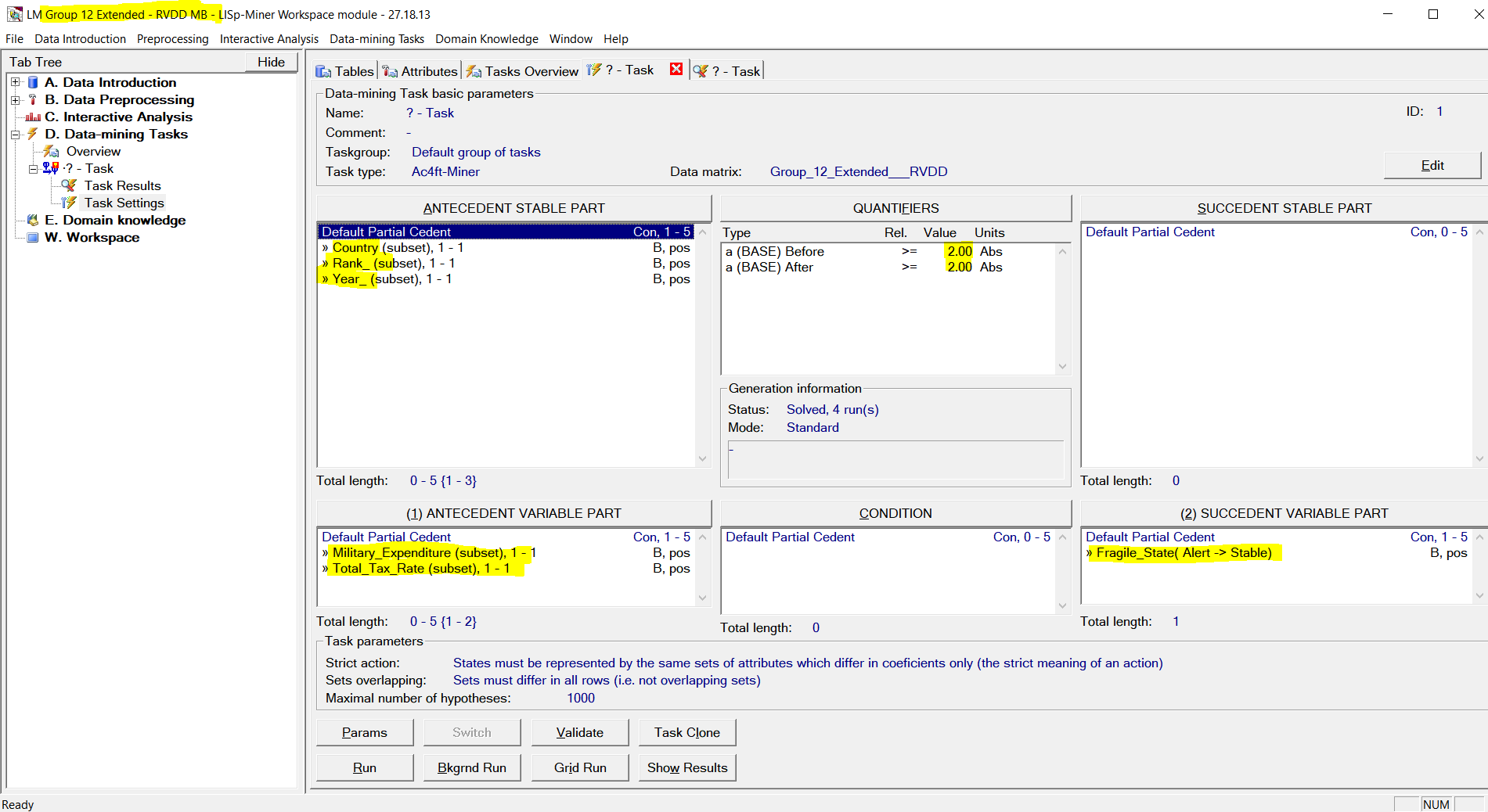
Categories



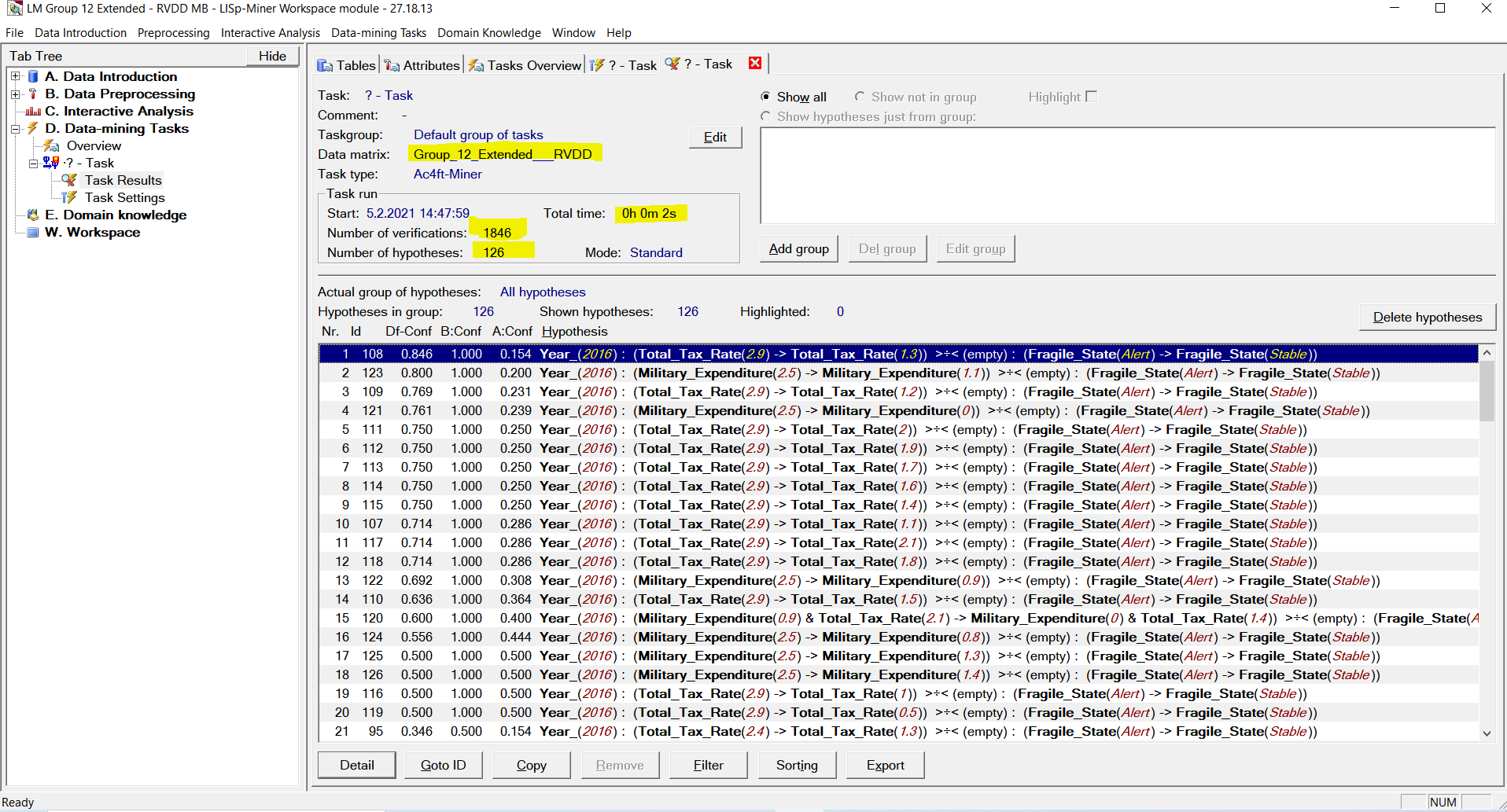
Selected Attributes



Task Overview

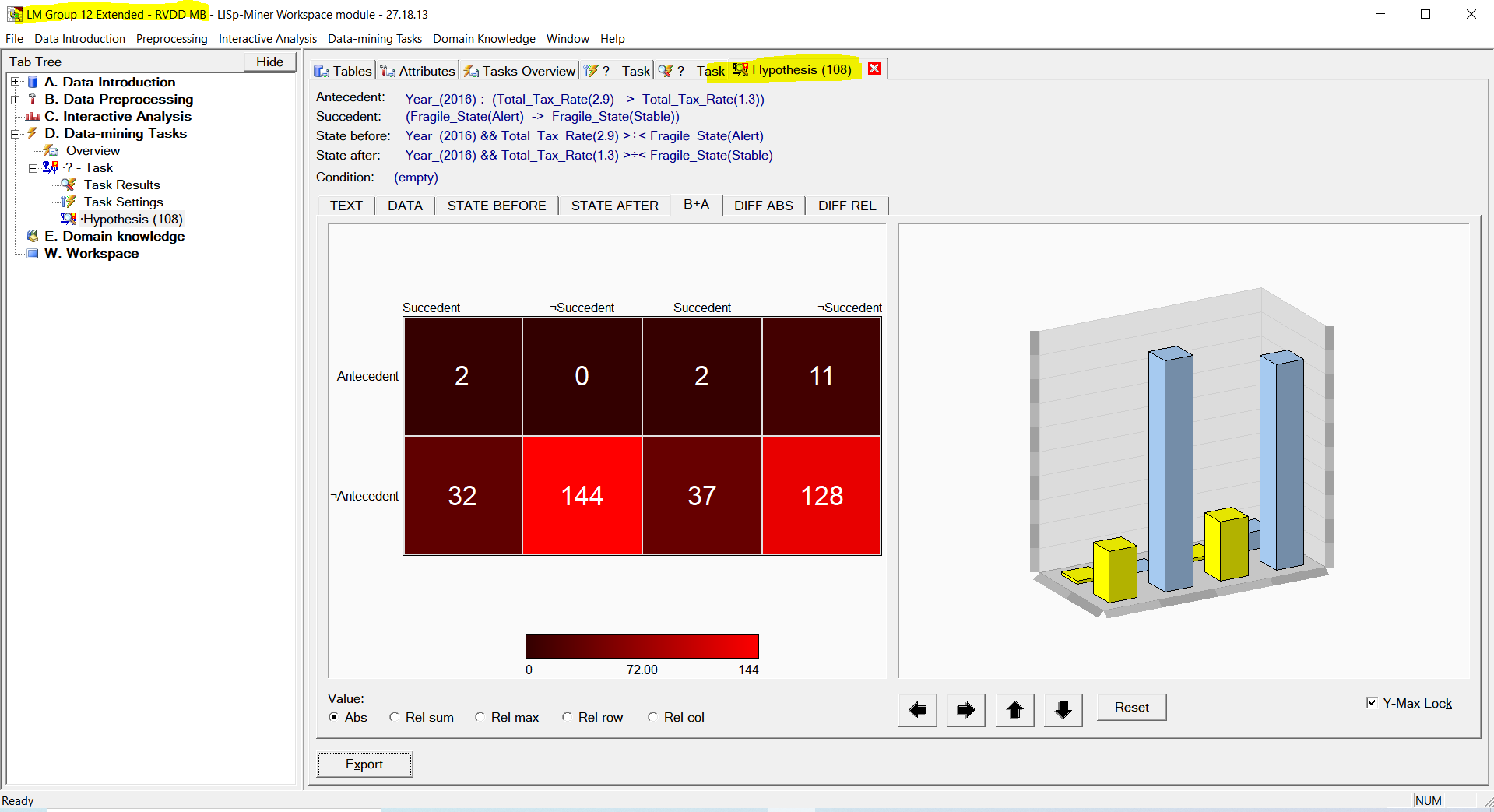


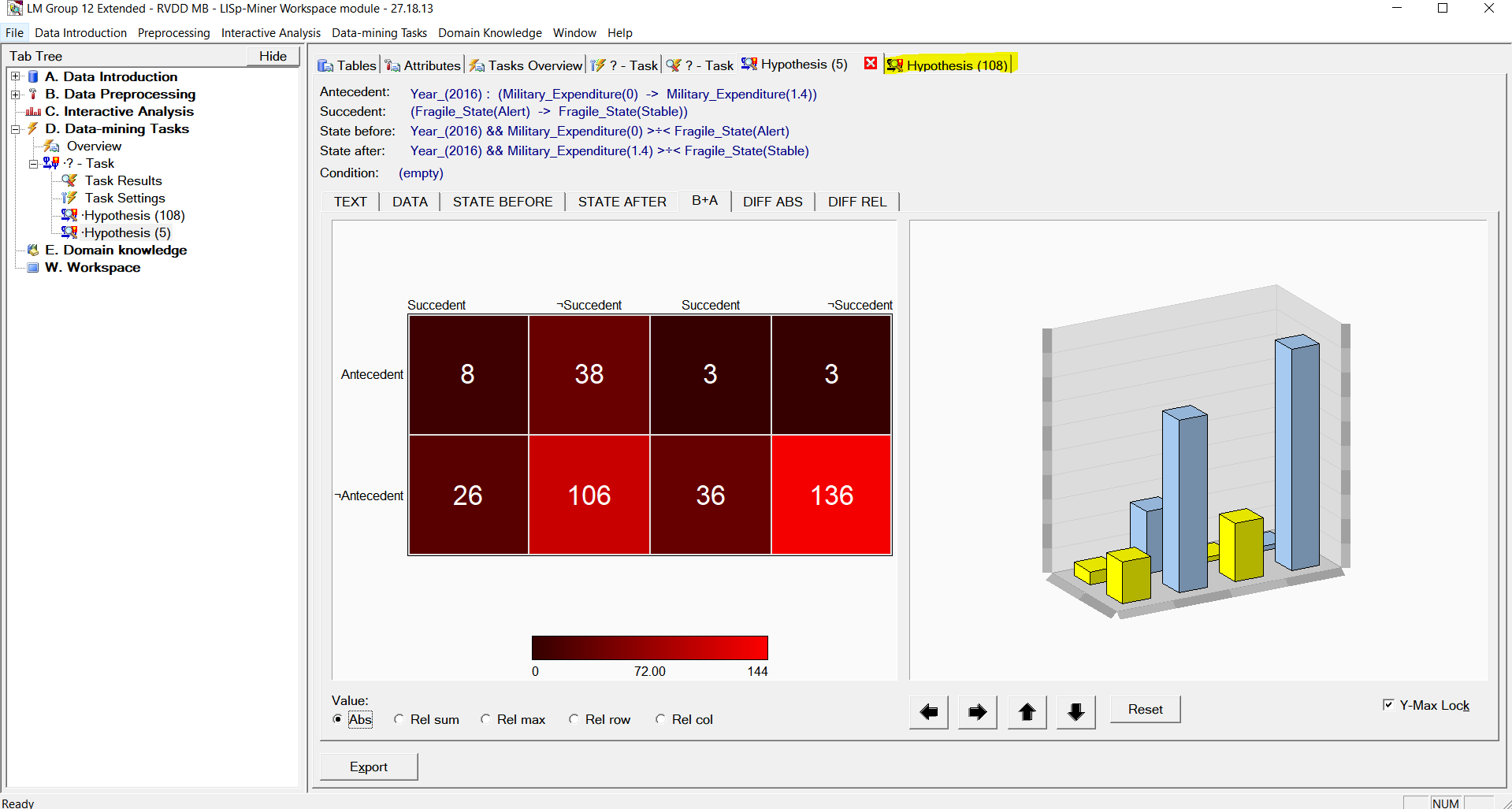
Parameter Settings

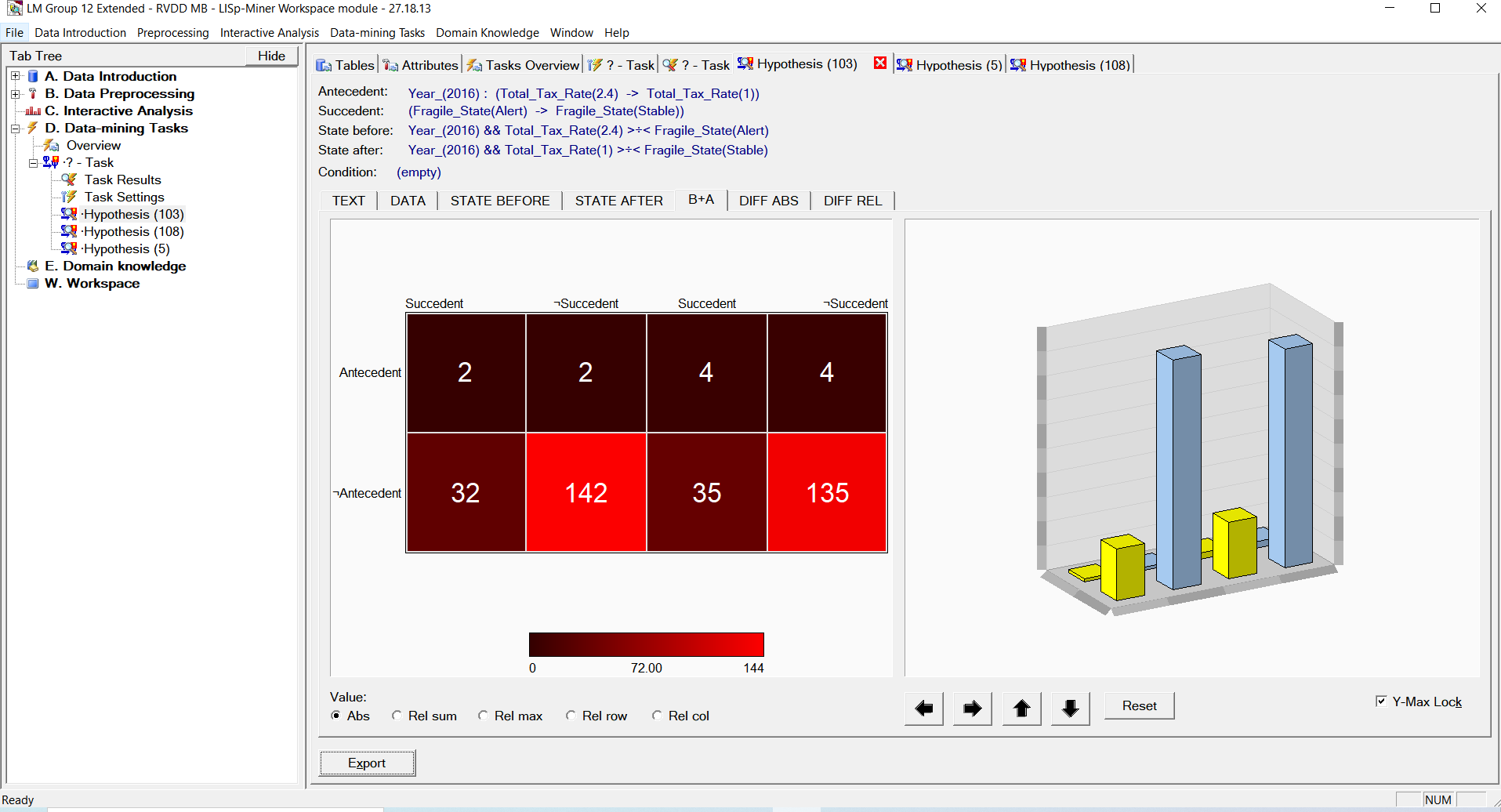


Task Results

Few Hypothesis Screenshots







Stable Attributes: Country, Year, Rank

Flexible Attributes: Military\_Expenditure, Total\_Tax\_Rate

Decision Attributes: Fragile State (Alert -> Stable)

**Fragile Attributes Used:**

* Military Expenditure: This attribute affects fragility in two ways, a higher index implies more fragility since a higher spending on military indicates more volatility. An increase in the value however may lead to lesser fragility since a nation prospers with better security
* Total\_Tax\_Rate: A change in this flexibility attribute may indicate progress in a sense that lower tax rates will lead to happier customers and lesser fragility

**Action Rules and Inferences:**

**Decision Attribute:** Fragile State

**Stable Attributes:** Country, rank and year

**Flexible attributes:** Military Expenditure, Total Tax Rate

**Action Rules: (Boundary values)**

1. (Military\_Expenditure(0) -> Military\_Expenditure(0.8)) >÷< (Fragile\_State(Alert) -> Fragile\_State(Stable))
2. (Military\_Expenditure(2.5) -> Military\_Expenditure(1.4)) >÷< (Fragile\_State(Alert) -> Fragile\_State(Stable))
3. (Total\_Tax\_Rate(1.3) -> Total\_Tax\_Rate(0.5)) >÷< (Fragile\_State(Alert) -> Fragile\_State(Stable))
4. (Total\_Tax\_Rate(2.9) -> Total\_Tax\_Rate(2.1)) >÷< (Fragile\_State(Alert) -> Fragile\_State(Stable))

**Inferences:**

1. It means higher security as military expenditure is slightly increased, resulting in a less fragile state.

2. Reducing the overall tax rate would mean less angry customers, and healthier companies, too, would help people in general and reduce fragility.

**Data Sources and Methods of Extraction:**

* To add additional features to our data sources, the FSI website and the World Bank website have been referenced and forked.
* The World Bank data website has World Development Indicator repositories that provide data for most countries around the world and for a wide variety of years.
* The data for the Countries in the world, World development indicators and wide range of years, were fetched from repositories of World banks Data.
* The data will now be a (comma separated delimited) CSV file.
* The downloaded data was not normalized in the beginning with respect to our range.
* Hence, we normalized this data to a range of 0-10 in Microsoft Excel
* This newly formatted data is then added to the datasets from the FSI website.

**Conclusion:**

* Information was collected through the web from different sources and information was pre-processed, discrete and grouped using WEKA.
* Activity rules using Lisp Miner were produced and investigated.
* In this way, recommendations for intervention can be used as a tool to assess a nation's condition and to take critical corrective steps to improve a nation's state.

**References:**

* <https://datamarket.com/data/set/14s4/military-expenditure-of-gdp>
* <https://en.wikipedia.org/wiki/Fragile_States_Index>
* <http://databank.worldbank.org/data/home.aspx>
* <http://fsi.fundforpeace.org/>
* <http://hdr.undp.org/en>
* [http://mecometer.com/infographic/switzerland/business-environment-taxes/](http://mecometer.com/infographic/switzerland/business-environment-taxes/%20)
* <http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/BayesNet.html>
* <http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/NaiveBayes.html>
* <http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/JRip.html>
* <https://data.worldbank.org>