**Logo, company name

Description automatically generated**

ITCS - 6162

KNOWLEDGE DISCOVERY IN DATABASES

PROJECT REPORT

**CLASSIFICATION AND EXTRACTION OF ACTION RULES ON THE FSI (FRAGILE STATE INDEX) DATA**

Professor – Dr. ZBIGNIEW W. RAS

Teaching Assistant 1 – Rishab Semlani

Teaching Assistant 2 - Sapna Pareek

By:

Rajat Maheshwari(801255895)

Aditeya Nanda(801211280)

Roop Sai Namala (801255251)

Pratyusha Tummuri(801254220)

**INDEX**

1. **PROJECT & PROBLEM DESCRIPTION**
   1. **FSI (Fragile State Index)**
2. **DATA – ORIGINAL & EXTENDED**
   1. **Cohesion Indicators**
   2. **Economic Indicators**
   3. **Political Indicators**
   4. **Social Indicators**
   5. **Cross-Cutting Indicators**
   6. **Extend Features**
3. **CLASSIFICATION**
   1. **Preprocessing**
   2. **Visualization**
   3. **Steps to load the dataset onto Weka**
   4. **Classifiers Used**
   5. **Results of Classifiers on both Original and Extended Dataset**
4. **ACTION RULE MINING**
   1. **Steps followed to achieve results on Lisp Miner**
   2. **Results of Lisp Miner**
5. **CONCLUSION**
6. **REFERENCES**

**PROJECT & PROBLEM DESCRIPTION**

The problem statement considered for setting up the project is: To analyze the features that affect a country’s FSI score and report the analysis to help in the lowering of FSI of a country. Hence, we have found a way to perform such an analysis.

This project is mainly divided into two parts:

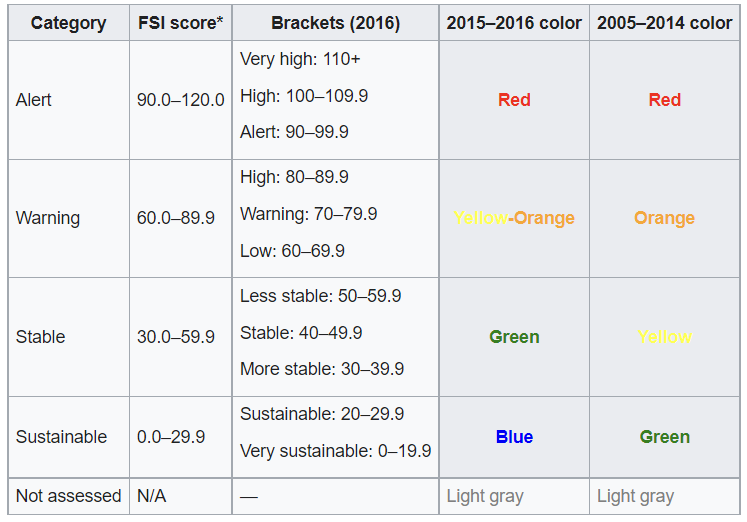
1. Obtaining the dataset, adding the external features to the original dataset, and performing classification on the extended dataset to check if the classification performance changes.
2. Mining the action rules on the extended dataset.

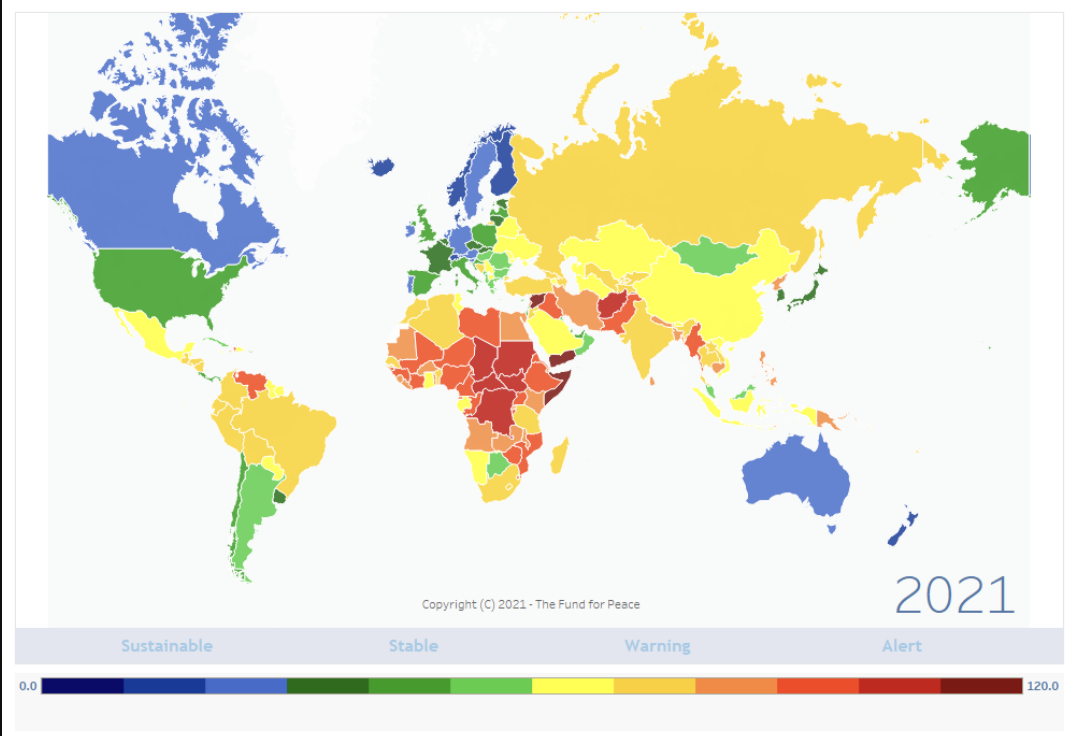
After we mine the action rules, we would be able to determine how a country is supposed to proceed further to lower its FSI score.

**FSI (Fragile State Index):**

The Fragile State Index was a report that is published every year to determine if a country is susceptible to collapse. This annual report was in practice from the year 2005, initiated by the United States. Now, this analysis is done by an organization named The New Humanitarian. There are about 179 countries involved in this report. A few countries are no longer considered in this analysis.

There are 12 major indicators that the organization uses to prepare the report of all the countries that are involved. Every indicator has a value between 0 – 10 where 0 is the best value and 10 is the worst value of an indicator. Since there are 12 attributes/indicators, the sum of all values of indicators is carried out. If the value of the sum is between 90 to 120, then the country is in an Alert state and at the risk of collapse. Similarly, 60-90 means that the country is in Warning state, 30-60 means Stable state, and 0-30 mean Sustainable state.





**DATA – ORIGINAL & EXTENDED**

The data is in the format of a CSV file. There are 12 indicators present and are valued from 0-10. The indicators are described as follows:

**COHESION INDICATORS:**

* **C1: Security Apparatus –** This indicator tells us about how much security threat the country is prone to. The security threats include bombings, wars, etc… Apart from the external threat, it also considers the internal threats that are plausible.
* **C2:** **Factionalized Elites –** This indicator considers if a country is divided based on caste, creed, race, religion, etc… The usage of these factors by the political elites is also considered.
* **C3:** **Group Grievance** – This indicator describes the division of multiple groups in the society based on various political and social features.

**ECONOMIC INDICATORS:**

* **E1: Economic Decline –** This indicator indicates how and why a country’s economy falls and the factors related to the fall. Some of the features that are considered are productivity, rate of unemployment, inflation, etc...
* **E2: Uneven Economic Development –** This indicator values the country’s inequality in the economy based on the structural inequality of the factors such as poverty, race, caste, etc…
* **E3: Human Flight and Brain Drain –** This indicator represents the number of people who migrate from their country to other countries for a better living. It could be the migration of the middle class or the migration of the elites. This causes a lack of professional and skillful labor.

**POLITICAL INDICATORS:**

* **P1: State Legitimacy –** This indicator tells us how open and communicative the ruling government is with the citizens of the country. Some factors like the surveys from the people for the satisfaction or happiness of the government are considered in this process.
* **P2: Public Services –** This indicator helps in determining the feasibility, the expenses, and the availability of all the services to the citizens provided by the government.
* **P3: Human Rights and Rule of Law** – It shows the value of how much the rules & regulations, the law is followed by the government and also the public to ensure proper development and justice.

**SOCIAL INDICATORS**:

* **S1: Demographic Pressures –** This indicator measures the value of pressure that is imposed on the country by the population for the utilities that are required to fill the everyday necessities.
* **S2:** **Refugees and IDPs –** This indicator measures the values of pressure imposed on the country by the people who migrate from or away from a country.

**CROSS-CUTTING INDICATORS:**

* **X1: External Intervention –** This indicator mainly focuses on the external factors that affect the country. Most of the factors in the indicators are related to other countries that are both covert and overt.

**EXTENDED FEATURES / INDICATORS:**

Considering the original data from the FSI-2020 file, we have added 6 new features to improve the classification and help us in better identification of action rules that help the country to develop itself from collapsing. The major source for the data was the website [www.data.worldbank.org](http://www.data.worldbank.org). The data extraction from the sources was done manually i.e. each value was entered for every country and was pre-processed later.

The 6 features added were:

* **Corruption Index**: The Corruption Index of a country represents the score a country gets when referring to the perceived levels of corruption that was done by the government of that country. These scores are given on a scale of 0 to 100 where 0 is the high level of corruption and 100 is the low level of corruption. We have considered this as a factor because the country can collapse if the government is extremely corrupt and there won’t be any development in terms of economy, infrastructure, etc… So, we have scaled the values on the range of 0 to 10 and later subtracted the obtained value from 10 so that 10 means a higher value of corruption and 0 means a low value of corruption.
* **Terrorism Index**: The Terrorism Index, also known as the Global Terrorism Index is an indicator that is used to represent the terrorism activity that happens in every country. There are a few features that represent a value of the Terrorism Index such as, Number of terror activities, number of fatalities due to terrorist activities, etc... We have considered this indicator because, if a country is prone to more terrorist attacks or if a country is known for terrorist activities, then there won’t be proper investments, no proper resources will be given for development, etc… This could put a country at the risk of collapse. The values are scaled between 0 to 10 where 10 means the worst countries and 0 is peaceful countries.
* **Gini Index**: The Gini Index is one of the most popular metric used by the World Bank to consider the financial and economical situation of a country. Consider an equal distribution of income in the economy amongst all classes of people in the country. This index value represents the deviation of the income of people from the equal distribution. The range of this index is from 0 to 100 where 0 means proper equality and 100 means extreme inequality. We’ve scaled this value to 0-10. We’ve chosen the Gini Index because, the higher the Gini index, the greater the deviation from the equality. This means that only one class of people would be contributing to the economy while others have little to do with it. These countries would get poorer and poorer if this inequality is not balanced.
* **Trade in Services**: Trade in Services can be defined as the purchase and delivery of a specific service between any producer and any consumer. This happens on a large scale between countries where the Producer and Consumer are 2 different countries. The indicator could be better while identifying the Action Rules because, if a country is only a consumer of a service and no production happens, then there would be a rise in expenditure causing fluctuations in the economy and vice-versa. If a country offers more services, it means that the country has no reason to collapse. On the other hand, if a country consumes more than it offers, then the country might collapse.
* **Inflation:** In the terms of Economy, inflation refers to a linear increase in the prices of goods and services in a country or a province. This means that one will buy fewer goods for a larger value of money. Inflation occurs if a country is falling short of resources. So, to meet the needs of the country's economy, the process of inflation is brought into action. If a country is under inflation, then it is an indicator that the specific country is on the edge of a collapse. Hence, this could be a good indicator and we have used it to determine the FSI state of a country.
* **Urban Population:** The Urban Population refers to the percentage of the population that lives in the developed parts i.e. urban parts of the country. If the percentage is high, then it means that most parts of the country are properly developed, the revenue being generated would be high, and there is a smaller risk for collapse. If the percentage is low, then there is a higher risk for collapse as the places are under-developed and lesser revenue is generated.

**CLASSIFICATION**

**PREPROCESSING**

To get perfect results in a classification task, the data that we use have to be perfect. So, to ensure that the data is perfect, we have to preprocess it before classification. Some of the preprocessing tasks we have done on the dataset are:

* After the addition of features, we have computed the sum and scaled the sum value to the range of 0-120.
* For some indicators, high value meant sustainable countries. So, we have subtracted that value from 10 so that a lower value means sustainable countries as this is the pattern of the dataset.
* From the dataset, we have removed a column that said “Change from previous year” as this did not indicate any use for the task of classification.

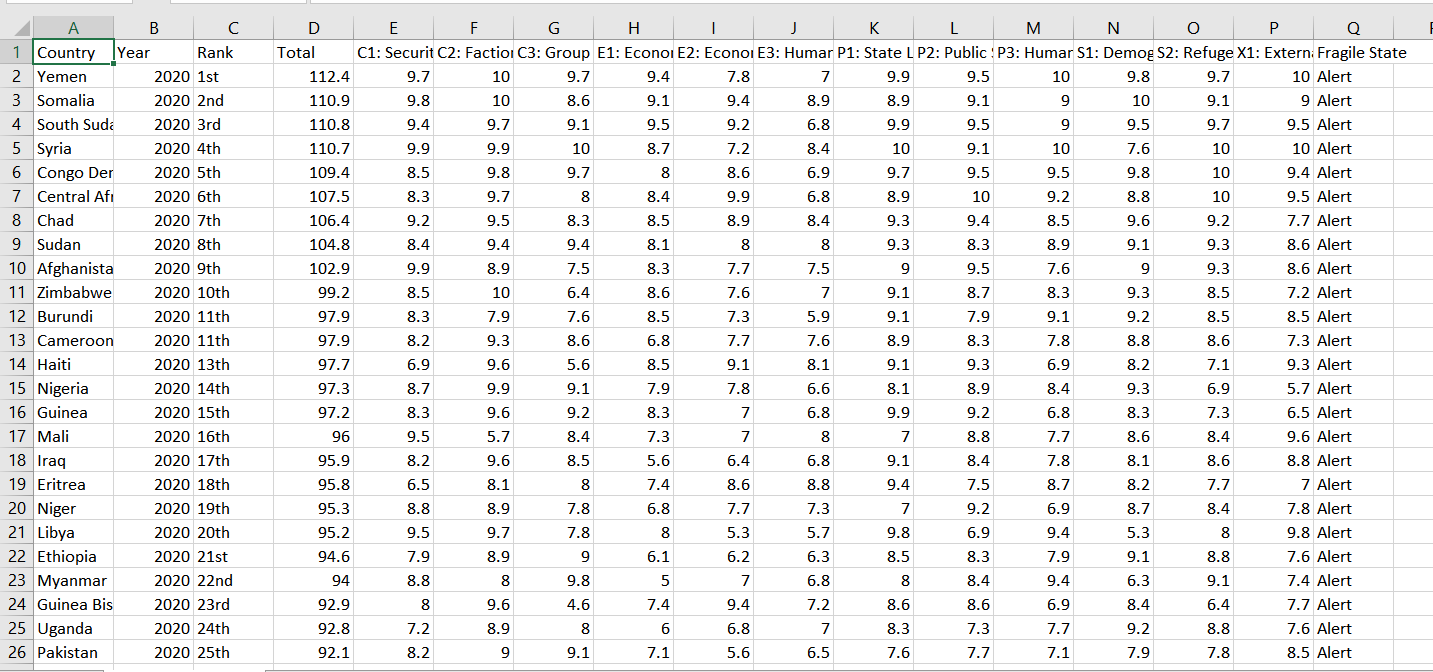
**VISUALIZATION**

Visualization of the data was important as we were able to ensure that most of the major indicators were of the same format of data. Here are some of the images from the visualization:

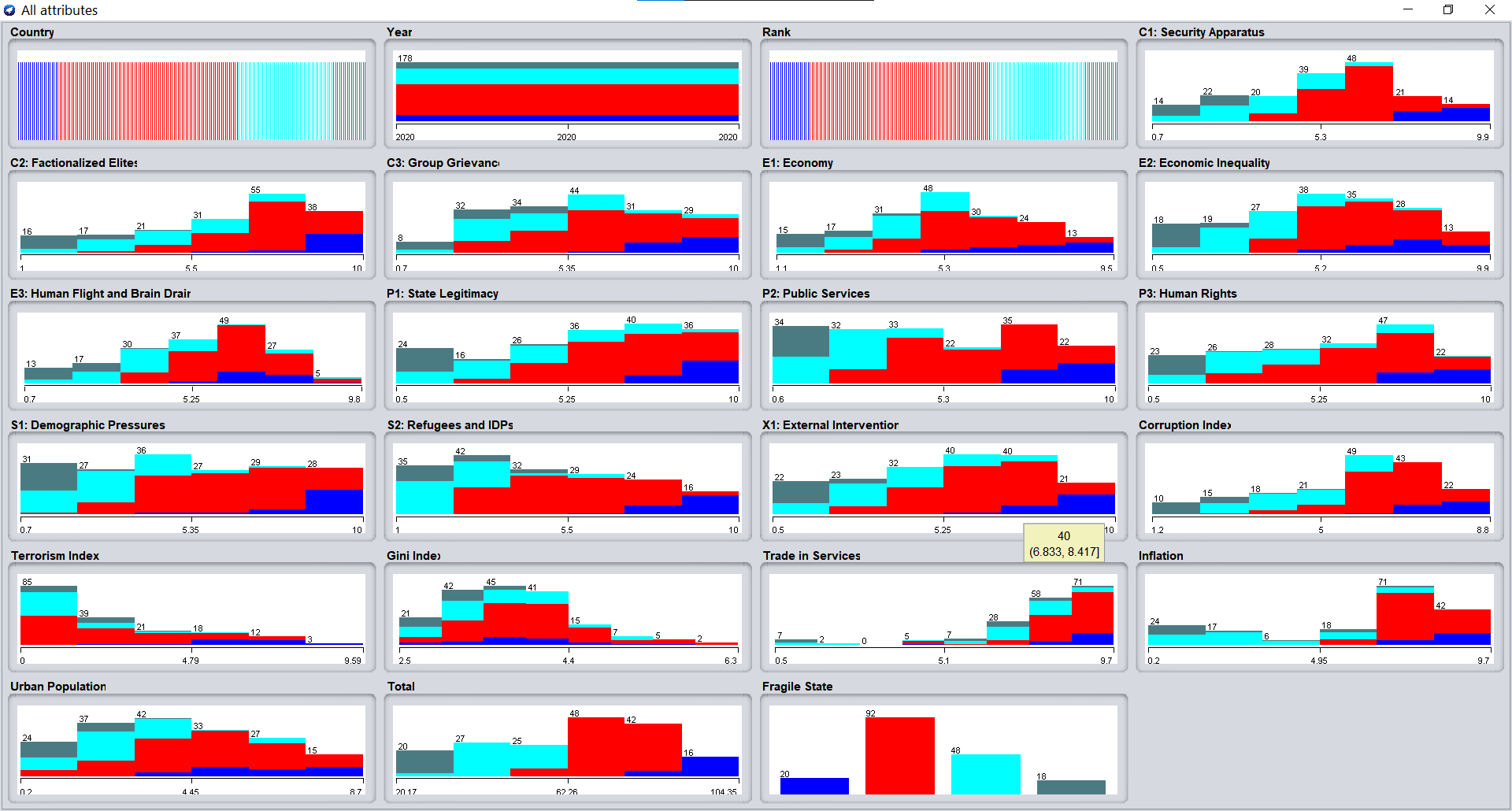
1. Visualizing the Original FSI-2020 Dataset

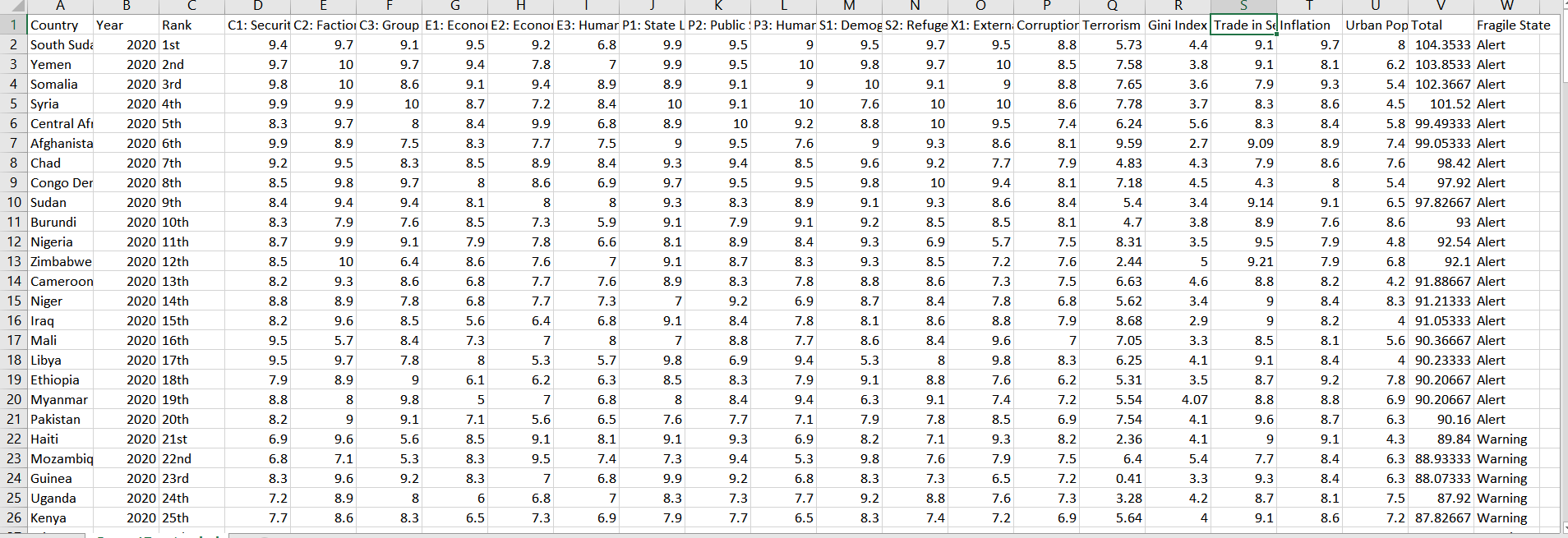
Timeline

Description automatically generated



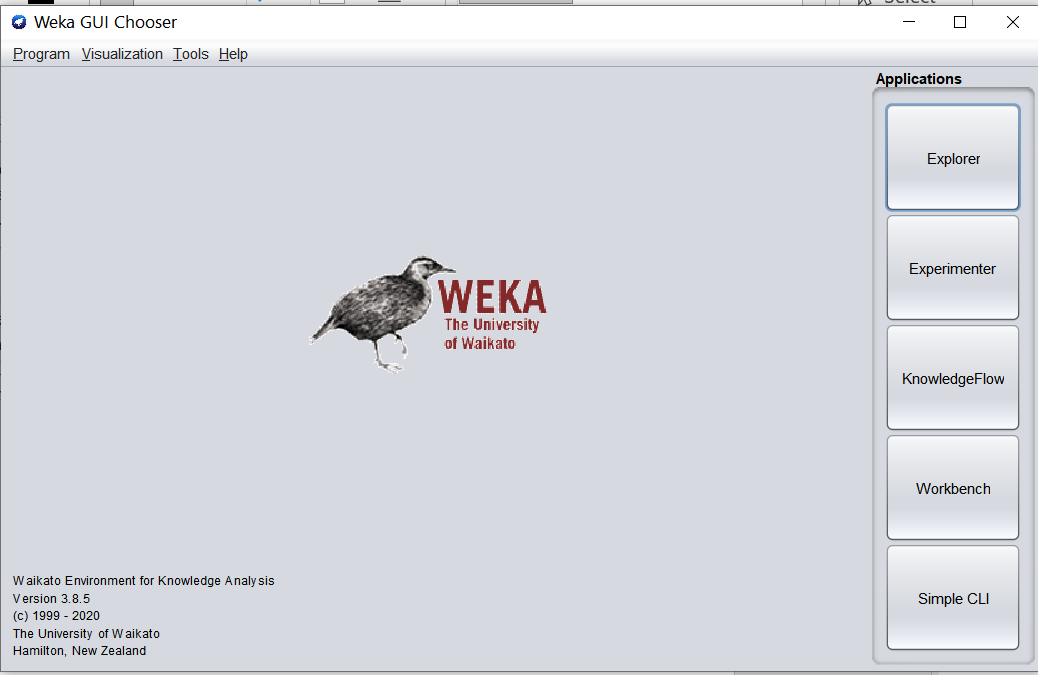
1. Visualizing the Extended Dataset:



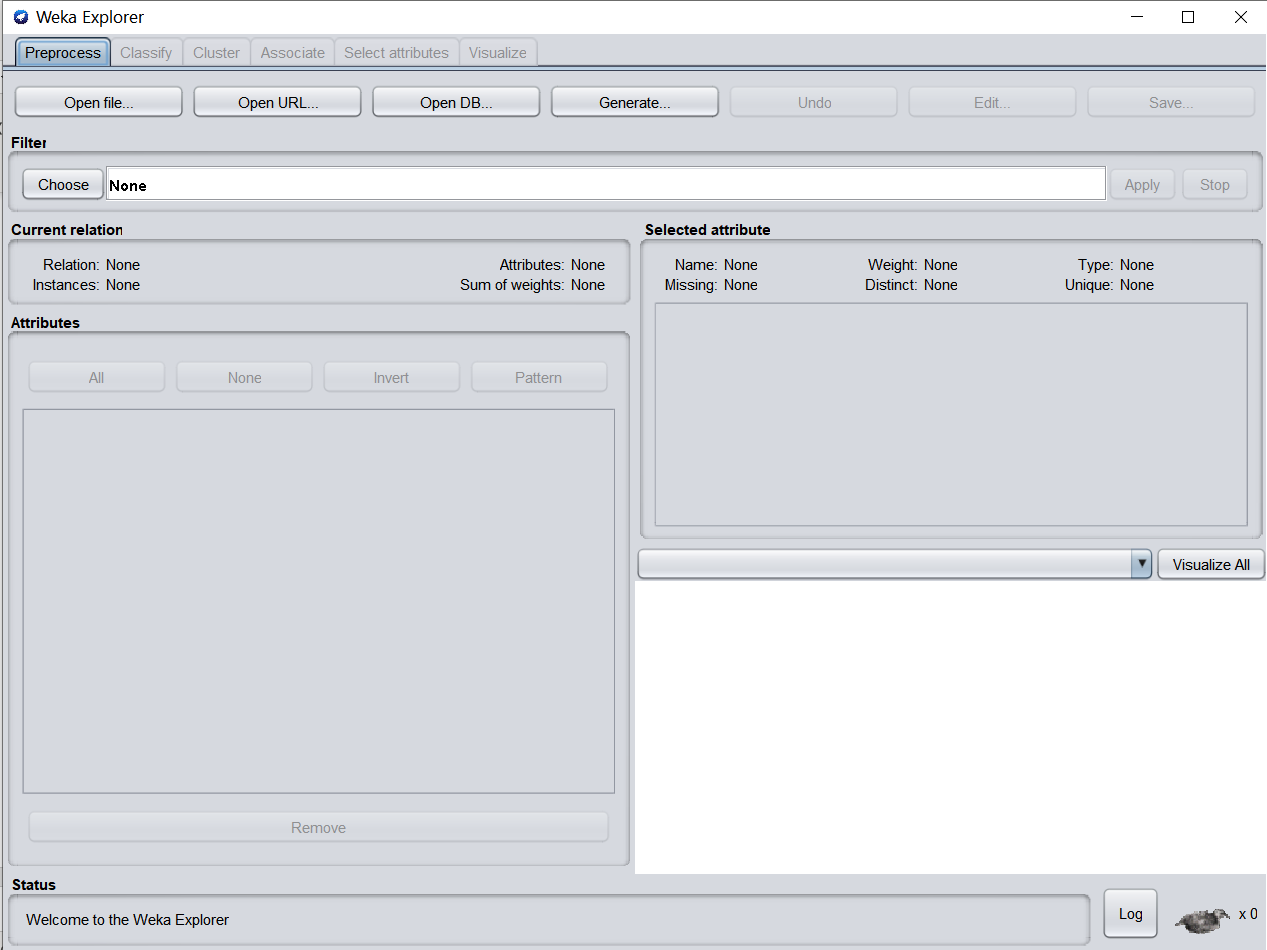


**STEPS TO LOAD THE DATASET ONTO WEKA**

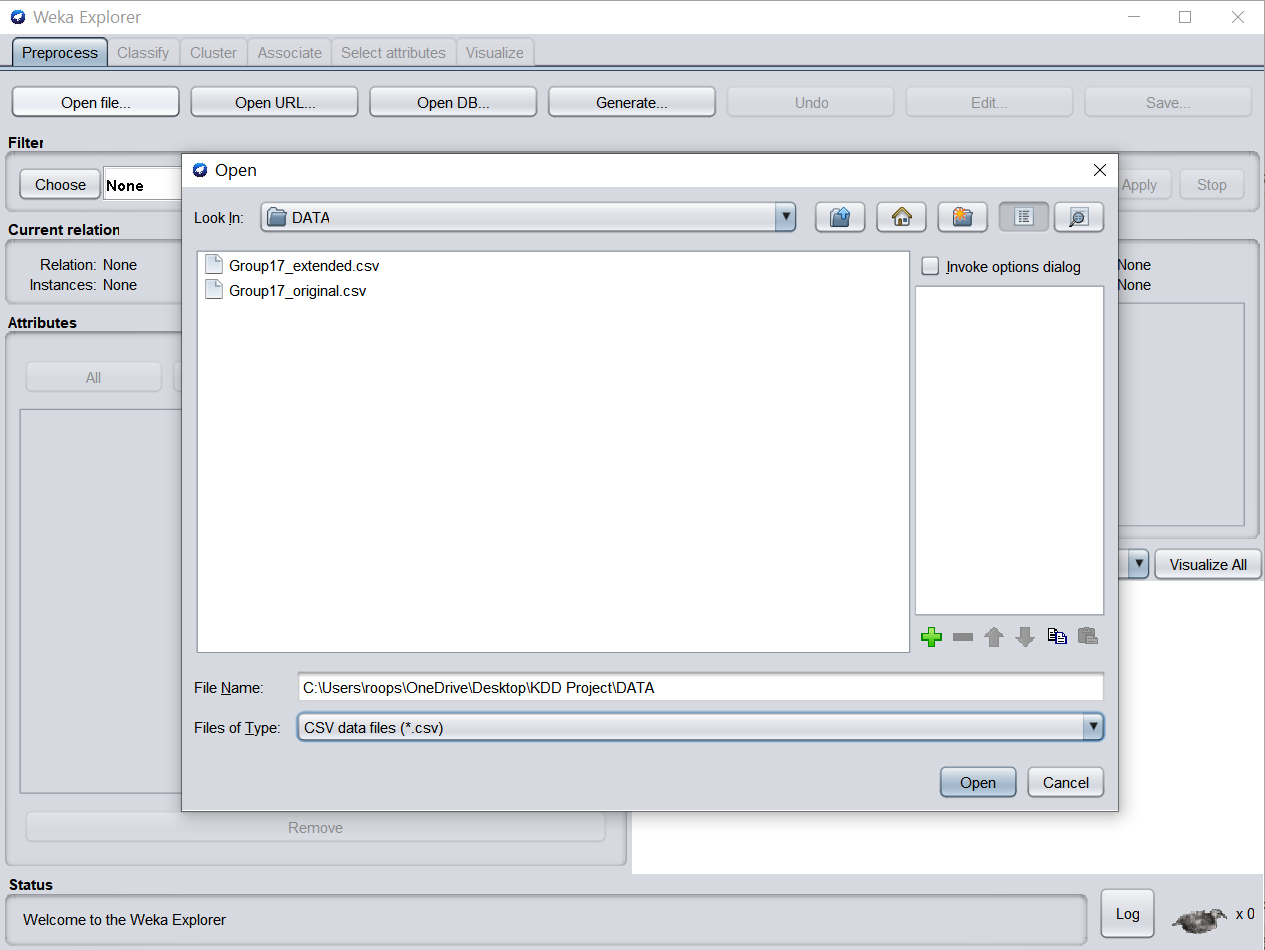
1)Install and Open Weka. After opening, click on the explorer tab on the right side of the tab.



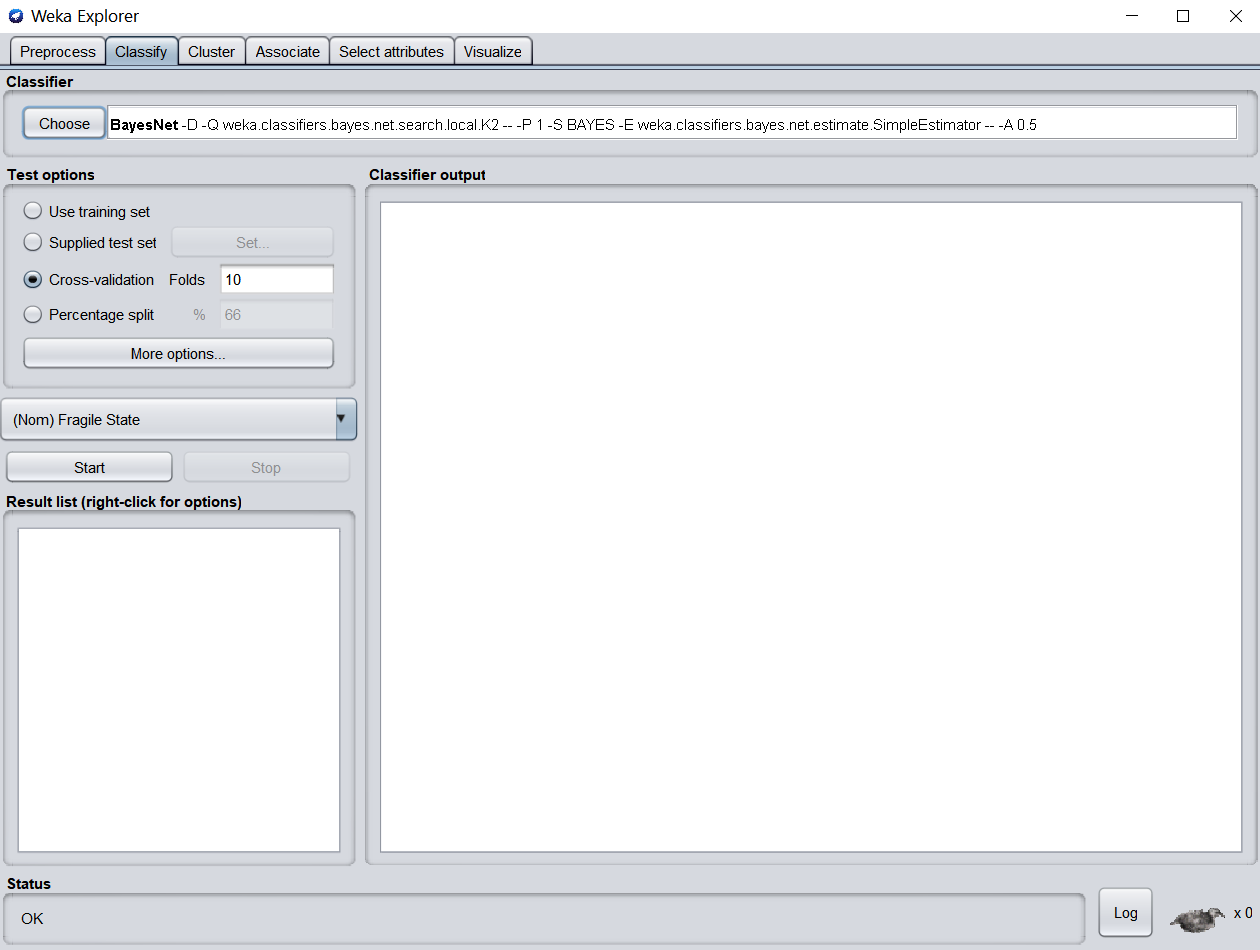
2)After pressing the explorer, a window pops up. Click on the open file button on the top left of the window



3)Navigate to the location of your file and choose the file in .csv format.



4)Open the file, press on classify button on the top, choose the classifier, and press start



**CLASSIFIERS USED**

1. **Bayesian Network Classifier:**

Bayesian Network Classifier is a classifier that is based on Bayes' Theorem and assumes strong (Naive) independence assumptions. The underlying probability model would be better described as a "independent feature model." A naïve Bayes classifier, in simple terms, asserts that the presence (or absence) of one feature of a class is unrelated to the presence (or lack) of any other feature.

1. **J48 Classification algorithm:**

The J48 algorithm is one of the most extensively used machine learning algorithms for categorizing and continuously examining data. It is commonly used to categorize data in a variety of sectors, such as evaluating clinical data for the diagnosis of coronary heart disease, classifying E-governance data, and so on. Using the concept of information entropy, this method creates decision trees from a set of training data. The training data consists of a collection of already identified samples S=s1, s2,... Each sample si is represented by a p-dimensional vector (x1, I x2, I xp, I where the xj reflects the relevant sample's attribute values or characteristics, as well as the class in which the sample belongs. The ideal attribute to split on for the best classification accuracy is the attribute with the most information.

1. **LMT Classifier:**

A logistic model tree (LMT) is a classification model that combines logistic regression with decision tree learning with an accompanying supervised training technique. Logistic model trees are built on the concept of a model tree, which is a decision tree with linear regression models at the leaves to build a piecewise linear regression model. The method can handle binary and multi-class target variables, numeric and nominal properties, and missing values, among other things.

**RESULTS OF THE CLASSIFIERS ON BOTH ORIGINAL & EXTENDED DATASET**

We’ve used the WEKA 3.8.5 software for classification purposes. This software is available to all and open to access.

1. **Results given by the BayesNet classifier on the original dataset:**

=== 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: Group17\_original

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

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

Year(1): Fragile State

Rank(162): Fragile State

Total(4): Fragile State

C1: Security Apparatus(4): Fragile State

C2: Factionalized Elites(5): Fragile State

C3: Group Grievance(3): Fragile State

E1: Economy(4): Fragile State

E2: Economic Inequality(4): Fragile State

E3: Human Flight and Brain Drain(4): Fragile State

P1: State Legitimacy(5): Fragile State

P2: Public Services(5): Fragile State

P3: Human Rights(4): Fragile State

S1: Demographic Pressures(4): Fragile State

S2: Refugees and IDPs(4): Fragile State

X1: External Intervention(4): Fragile State

Fragile State(4):

LogScore Bayes: -3979.223585397947

LogScore BDeu: -14809.38795998837

LogScore MDL: -11195.548589235394

LogScore ENTROPY: -7259.983982788541

LogScore AIC: -8778.98398278854

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 174 97.7528 %

Incorrectly Classified Instances 4 2.2472 %

Kappa statistic 0.9664

Mean absolute error 0.0131

Root mean squared error 0.0978

Relative absolute error 3.8899 %

Root relative squared error 23.8825 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

0.968 0.007 0.968 0.968 0.968 0.961 0.998 0.991 Alert

0.988 0.022 0.977 0.988 0.982 0.966 0.998 0.998 Warning

0.977 0.007 0.977 0.977 0.977 0.970 0.999 0.998 Stable

0.944 0.000 1.000 0.944 0.971 0.969 1.000 1.000 Sustainable

Weighted Avg. 0.978 0.013 0.978 0.978 0.977 0.966 0.998 0.997

=== Confusion Matrix ===

a b c d <-- classified as

30 1 0 0 | a = Alert

1 84 0 0 | b = Warning

0 1 43 0 | c = Stable

0 0 1 17 | d = Sustainable

Graphical user interface, application

Description automatically generated

1. **Results given by the BayesNet classifier on the Extended Dataset:**

=== 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: Group17\_extended

Instances: 178

Attributes: 23

Country

Year

Rank

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

Corruption Index

Terrorism Index

Gini Index

Trade in Services

Inflation

Urban Population

Total

Fragile State

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

Year(1): Fragile State

Rank(178): Fragile State

C1: Security Apparatus(5): Fragile State

C2: Factionalized Elites(4): Fragile State

C3: Group Grievance(4): Fragile State

E1: Economy(4): Fragile State

E2: Economic Inequality(4): Fragile State

E3: Human Flight and Brain Drain(4): Fragile State

P1: State Legitimacy(4): Fragile State

P2: Public Services(5): Fragile State

P3: Human Rights(4): Fragile State

S1: Demographic Pressures(4): Fragile State

S2: Refugees and IDPs(4): Fragile State

X1: External Intervention(4): Fragile State

Corruption Index(4): Fragile State

Terrorism Index(2): Fragile State

Gini Index(2): Fragile State

Trade in Services(2): Fragile State

Inflation(3): Fragile State

Urban Population(2): Fragile State

Total(4): Fragile State

Fragile State(4):

LogScore Bayes: -4681.86535836559

LogScore BDeu: -16169.105794174342

LogScore MDL: -12332.785280698672

LogScore ENTROPY: -8138.13149673722

LogScore AIC: -9757.13149673722

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 173 97.191 %

Incorrectly Classified Instances 5 2.809 %

Kappa statistic 0.9562

Mean absolute error 0.0157

Root mean squared error 0.1047

Relative absolute error 4.9068 %

Root relative squared error 26.2085 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

1.000 0.006 0.952 1.000 0.976 0.973 1.000 1.000 Alert

0.978 0.000 1.000 0.978 0.989 0.978 1.000 1.000 Warning

0.958 0.015 0.958 0.958 0.958 0.943 0.998 0.994 Stable

0.944 0.013 0.895 0.944 0.919 0.910 0.998 0.982 Sustainable

Weighted Avg. 0.972 0.006 0.973 0.972 0.972 0.961 0.999 0.997

=== Confusion Matrix ===

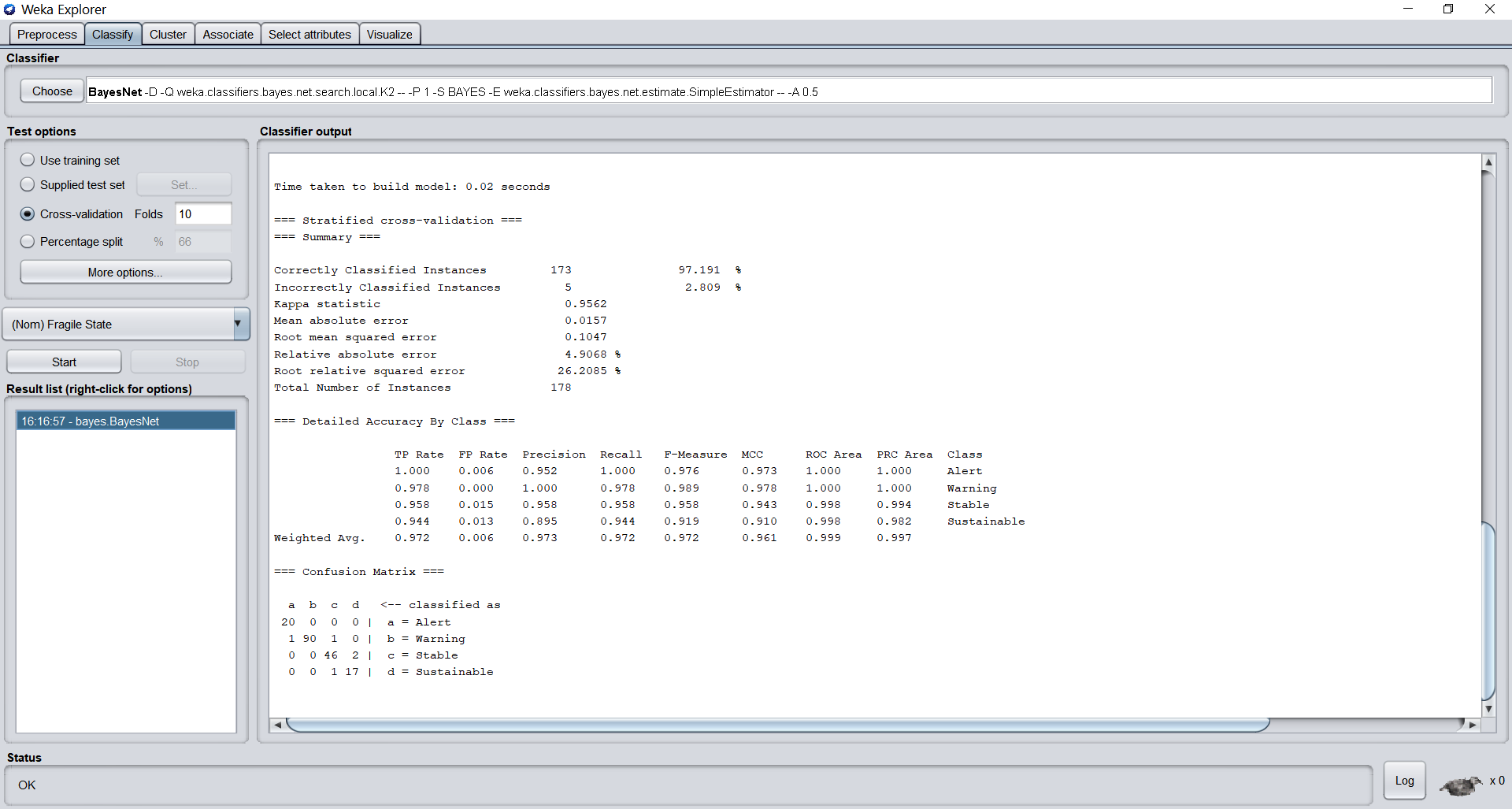
a b c d <-- classified as

20 0 0 0 | a = Alert

1 90 1 0 | b = Warning

0 0 46 2 | c = Stable

0 0 1 17 | d = Sustainable



1. **Results given by J48 classifier on the Original Dataset**:

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: Group17\_original

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

Test mode: 10-fold cross-validation

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

J48 pruned tree

------------------

Total <= 59.8

| Total <= 27.1: Sustainable (18.0)

| Total > 27.1: Stable (44.0)

Total > 59.8

| Total <= 89.7: Warning (85.0)

| Total > 89.7: Alert (31.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 175 98.3146 %

Incorrectly Classified Instances 3 1.6854 %

Kappa statistic 0.9748

Mean absolute error 0.0084

Root mean squared error 0.0918

Relative absolute error 2.5062 %

Root relative squared error 22.4152 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

1.000 0.007 0.969 1.000 0.984 0.981 0.997 0.969 Alert

0.988 0.011 0.988 0.988 0.988 0.977 0.989 0.982 Warning

0.977 0.007 0.977 0.977 0.977 0.970 0.985 0.961 Stable

0.944 0.000 1.000 0.944 0.971 0.969 0.972 0.950 Sustainable

Weighted Avg. 0.983 0.008 0.983 0.983 0.983 0.975 0.987 0.971

=== Confusion Matrix ===

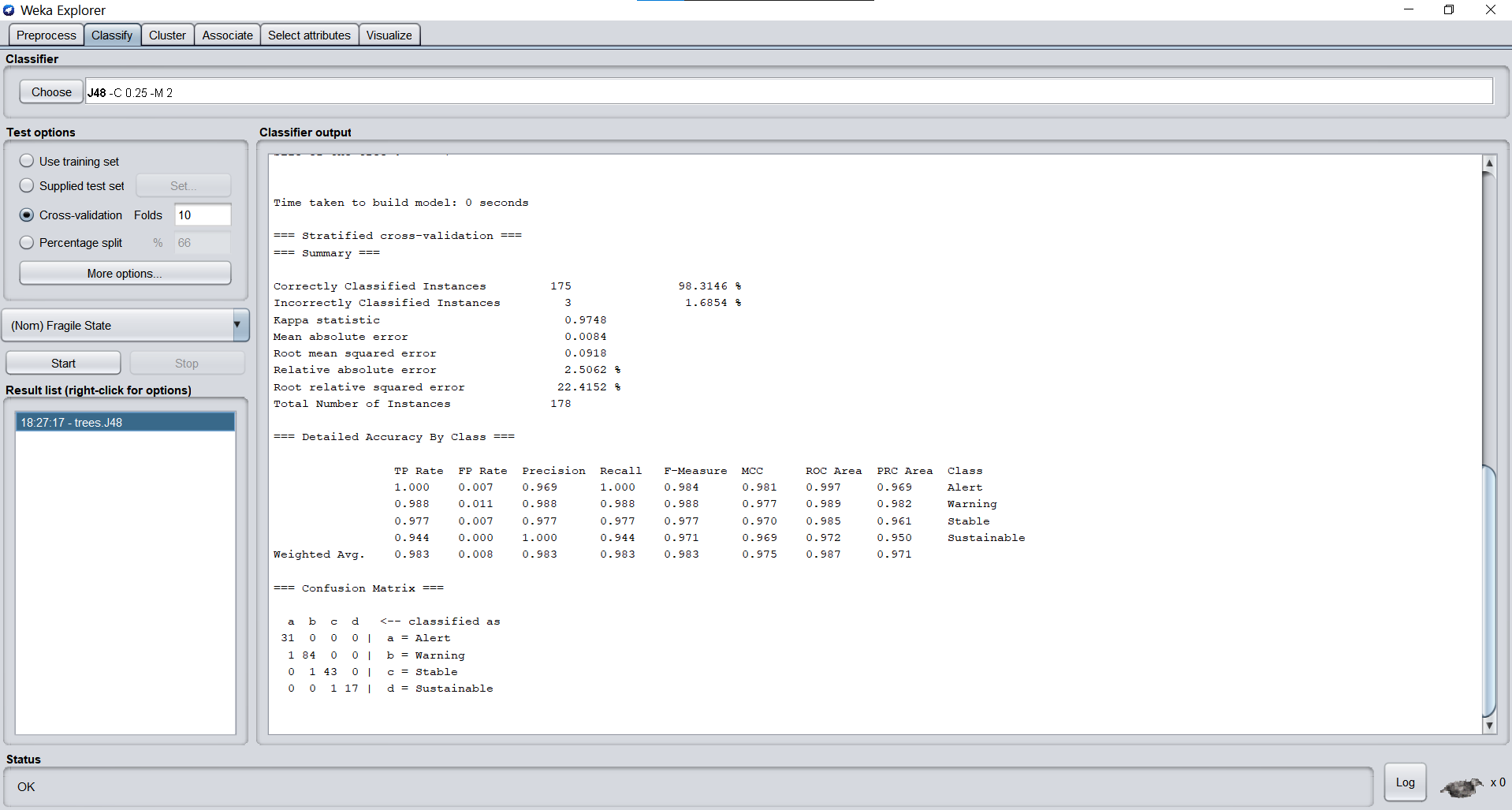
a b c d <-- classified as

31 0 0 0 | a = Alert

1 84 0 0 | b = Warning

0 1 43 0 | c = Stable

1. 0 1 17 | d = Sustainable



1. **Results given by J48 classifier on the Extended Dataset:**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: Group17\_extended

Instances: 178

Attributes: 23

Country

Year

Rank

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

Corruption Index

Terrorism Index

Gini Index

Trade in Services

Inflation

Urban Population

Total

Fragile State

Test mode: 10-fold cross-validation

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

J48 pruned tree

------------------

Total <= 59.933334

| Total <= 29.966667: Sustainable (18.0)

| Total > 29.966667: Stable (48.0)

Total > 59.933334

| Total <= 89.84: Warning (92.0)

| Total > 89.84: Alert (20.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 175 98.3146 %

Incorrectly Classified Instances 3 1.6854 %

Kappa statistic 0.9736

Mean absolute error 0.0084

Root mean squared error 0.0918

Relative absolute error 2.6326 %

Root relative squared error 22.9896 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

1.000 0.006 0.952 1.000 0.976 0.973 0.997 0.952 Alert

0.989 0.012 0.989 0.989 0.989 0.978 0.989 0.984 Warning

0.979 0.008 0.979 0.979 0.979 0.971 0.986 0.964 Stable

0.944 0.000 1.000 0.944 0.971 0.969 0.972 0.950 Sustainable

Weighted Avg. 0.983 0.009 0.983 0.983 0.983 0.974 0.987 0.972

=== Confusion Matrix ===

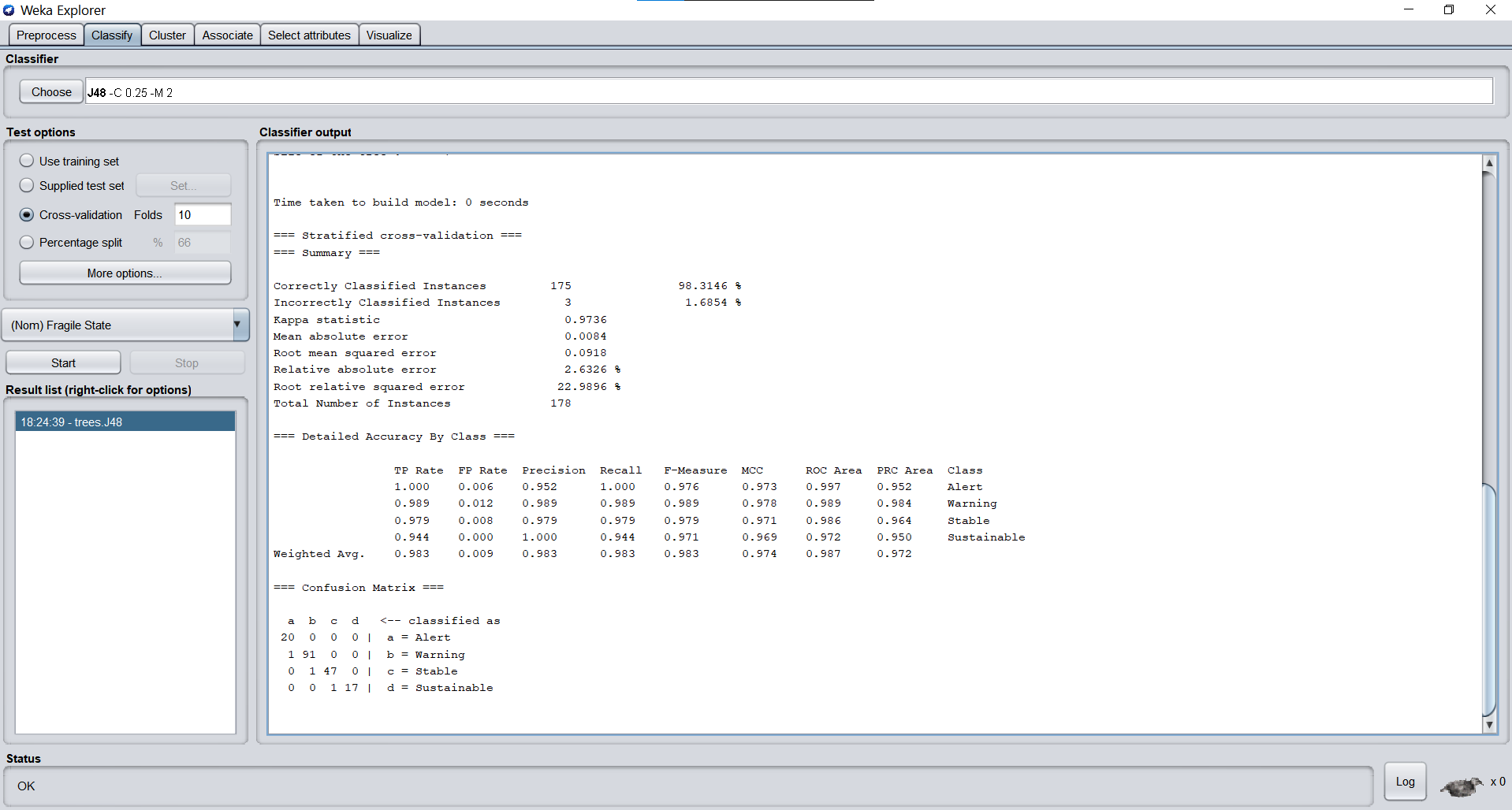
a b c d <-- classified as

20 0 0 0 | a = Alert

1 91 0 0 | b = Warning

0 1 47 0 | c = Stable

0 0 1 17 | d = Sustainable



1. **Results given by the LMT classifier on the Original Dataset:**

=== Run information ===

Scheme: weka.classifiers.trees.LMT -I -1 -M 15 -W 0.0

Relation: Group17\_original

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

Test mode: 10-fold cross-validation

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

Logistic model tree

------------------

: LM\_1:54/54 (178)

Number of Leaves : 1

Size of the Tree : 1

LM\_1:

Class Alert :

-734.5 +

[Country=Cote dIvoire] \* -22.74 +

[Country=Mauritania] \* -16.75 +

[Country=Angola] \* -8.99 +

[Total] \* 8.62 +

[C2: Factionalized Elites] \* -1.85 +

[P2: Public Services] \* 0.36 +

[S2: Refugees and IDPs] \* 0.38

Class Warning :

-54.67 +

[Country=Belarus] \* 1.06 +

[Country=Samoa] \* 0.92 +

[Country=Macedonia] \* 1.71 +

[Country=Belize] \* 0.89 +

[Country=Suriname] \* 4.69 +

[Country=Jamaica] \* 0.9 +

[Country=Kazakhstan] \* -2.32 +

[Country=Cuba] \* -0.91 +

[Country=Albania] \* -5.68 +

[Country=Malaysia] \* -1.49 +

[Country=Botswana] \* -0.79 +

[Country=Montenegro] \* -1.65 +

[Rank=110th] \* 4.36 +

[Total] \* 0.64 +

[C1: Security Apparatus] \* 0.63 +

[C2: Factionalized Elites] \* 0.29 +

[C3: Group Grievance] \* 0.17 +

[E1: Economy] \* 0.77 +

[E3: Human Flight and Brain Drain] \* 0.63 +

[P3: Human Rights] \* 0.11

Class Stable :

51.85 +

[Country=Belize] \* -1.87 +

[Country=Suriname] \* -0.9 +

[Country=Jamaica] \* -0.83 +

[Country=Kazakhstan] \* 4.58 +

[Country=Cuba] \* 4.04 +

[Country=Albania] \* 3.62 +

[Country=Malaysia] \* 1.67 +

[Country=Botswana] \* 4.96 +

[Country=Montenegro] \* 1.79 +

[Country=Grenada] \* 1.81 +

[Total] \* -0.69 +

[C1: Security Apparatus] \* -0.66 +

[C3: Group Grievance] \* -0.14 +

[E1: Economy] \* -0.77 +

[P2: Public Services] \* -1.04 +

[P3: Human Rights] \* -0.11 +

[S2: Refugees and IDPs] \* -0.4

Class Sustainable :

161.42 +

[Country=Malta] \* -1.5 +

[Country=Uruguay] \* -4.5 +

[Country=Japan] \* -6.02 +

[Country=South Korea] \* -4.5 +

[Country=France] \* -15.36 +

[Total] \* -4.62 +

[X1: External Intervention] \* -0.28

Time taken to build model: 0.71 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 171 96.0674 %

Incorrectly Classified Instances 7 3.9326 %

Kappa statistic 0.9413

Mean absolute error 0.023

Root mean squared error 0.1189

Relative absolute error 6.8306 %

Root relative squared error 29.0257 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

1.000 0.014 0.939 1.000 0.969 0.963 0.998 0.990 Alert

0.965 0.032 0.965 0.965 0.965 0.932 0.995 0.995 Warning

0.932 0.015 0.953 0.932 0.943 0.924 0.998 0.993 Stable

0.944 0.000 1.000 0.944 0.971 0.969 1.000 0.997 Sustainable

Weighted Avg. 0.961 0.021 0.961 0.961 0.961 0.939 0.997 0.994

=== Confusion Matrix ===

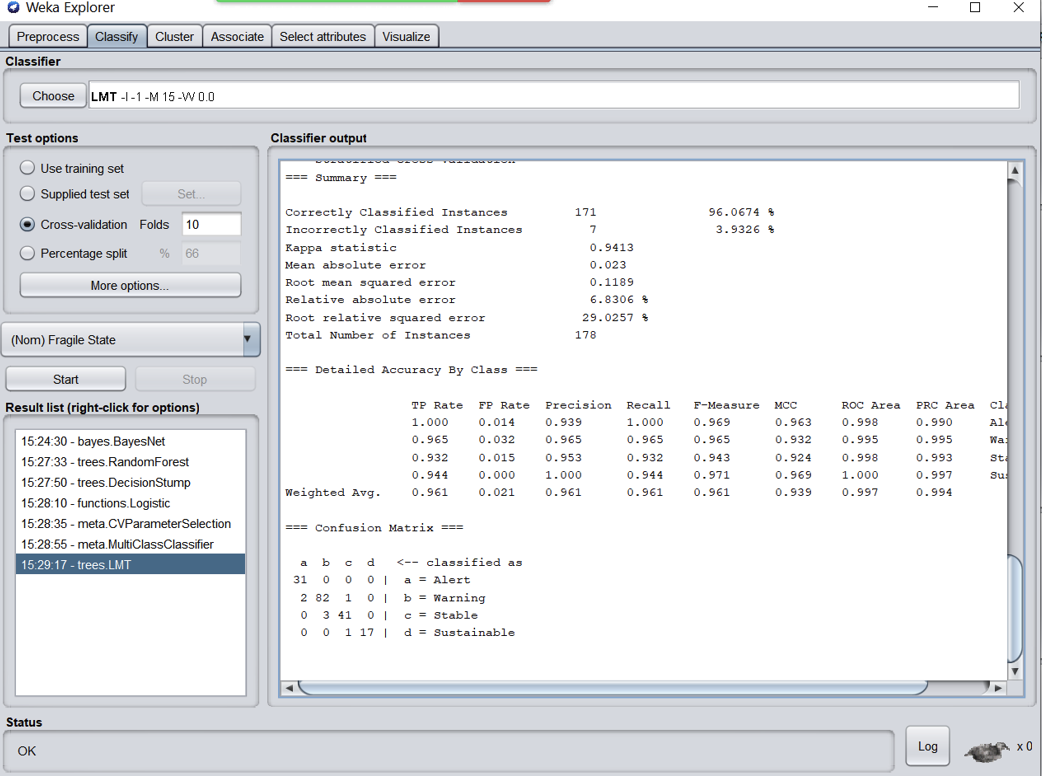
a b c d <-- classified as

31 0 0 0 | a = Alert

2 82 1 0 | b = Warning

0 3 41 0 | c = Stable

0 0 1 17 | d = Sustainable



1. **Results given by the LMT Classifier on the Extended Dataset:**

=== Run information ===

Scheme: weka.classifiers.trees.LMT -I -1 -M 15 -W 0.0

Relation: Group17\_extended

Instances: 178

Attributes: 23

Country

Year

Rank

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

Corruption Index

Terrorism Index

Gini Index

Trade in Services

Inflation

Urban Population

Total

Fragile State

Test mode: 10-fold cross-validation

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

Logistic model tree

------------------

: LM\_1:34/34 (178)

Number of Leaves : 1

Size of the Tree : 1

LM\_1:

Class Alert :

-710.95 +

[Country=Haiti] \* -4.49 +

[Country=Mozambique] \* -1.5 +

[C1: Security Apparatus] \* 0.72 +

[C3: Group Grievance] \* 0.9 +

[E2: Economic Inequality] \* -0.46 +

[P2: Public Services] \* 0.27 +

[S2: Refugees and IDPs] \* 0.38 +

[Terrorism Index] \* 0.3 +

[Total] \* 7.89

Class Warning :

-37.89 +

[Country=Saudi Arabia] \* 1.17 +

[Country=Vietnam] \* 1.2 +

[Country=Belize] \* 0.98 +

[Country=Dominican Republic] \* 0.89 +

[Country=Suriname] \* 1.78 +

[Country=Kazakhstan] \* -0.79 +

[Country=Malaysia] \* -2.24 +

[Country=Jamaica] \* -2.21 +

[Country=Grenada] \* -1.64 +

[Country=Trinidad and Tobago] \* -1.88 +

[Country=Panama] \* -2.25 +

[C2: Factionalized Elites] \* 0.18 +

[E2: Economic Inequality] \* 0.84 +

[E3: Human Flight and Brain Drain] \* 0.83 +

[Corruption Index] \* 0.7 +

[Terrorism Index] \* -0.13 +

[Gini Index] \* 0.39 +

[Trade in Services] \* 0.37 +

[Inflation] \* 0.56 +

[Total] \* 0.26

Class Stable :

30.61 +

[Country=Saudi Arabia] \* -1.09 +

[Country=Belize] \* -0.86 +

[Country=Suriname] \* -0.85 +

[Country=Kazakhstan] \* 1.98 +

[Country=Malaysia] \* 2.37 +

[Country=Jamaica] \* 3.17 +

[Country=Cuba] \* 3.07 +

[Country=Grenada] \* 1.77 +

[Country=Trinidad and Tobago] \* 1.91 +

[C2: Factionalized Elites] \* -0.17 +

[E3: Human Flight and Brain Drain] \* -0.42 +

[P2: Public Services] \* -0.56 +

[S2: Refugees and IDPs] \* -1.24 +

[Corruption Index] \* -0.27 +

[Terrorism Index] \* -0.38 +

[Inflation] \* -0.34 +

[Total] \* -0.3

Class Sustainable :

122.65 +

[Country=Japan] \* -2.98 +

[Country=Uruguay] \* -1.49 +

[Country=Slovenia] \* -13.52 +

[E2: Economic Inequality] \* -1.34 +

[Corruption Index] \* -0.38 +

[Trade in Services] \* -0.2 +

[Total] \* -3.33

Time taken to build model: 0.35 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 173 97.191 %

Incorrectly Classified Instances 5 2.809 %

Kappa statistic 0.9558

Mean absolute error 0.017

Root mean squared error 0.1161

Relative absolute error 5.3181 %

Root relative squared error 29.0631 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 Alert

1.000 0.012 0.989 1.000 0.995 0.989 1.000 1.000 Warning

0.938 0.015 0.957 0.938 0.947 0.928 0.992 0.970 Stable

0.889 0.013 0.889 0.889 0.889 0.876 0.997 0.980 Sustainable

Weighted Avg. 0.972 0.011 0.972 0.972 0.972 0.962 0.997 0.990

=== Confusion Matrix ===

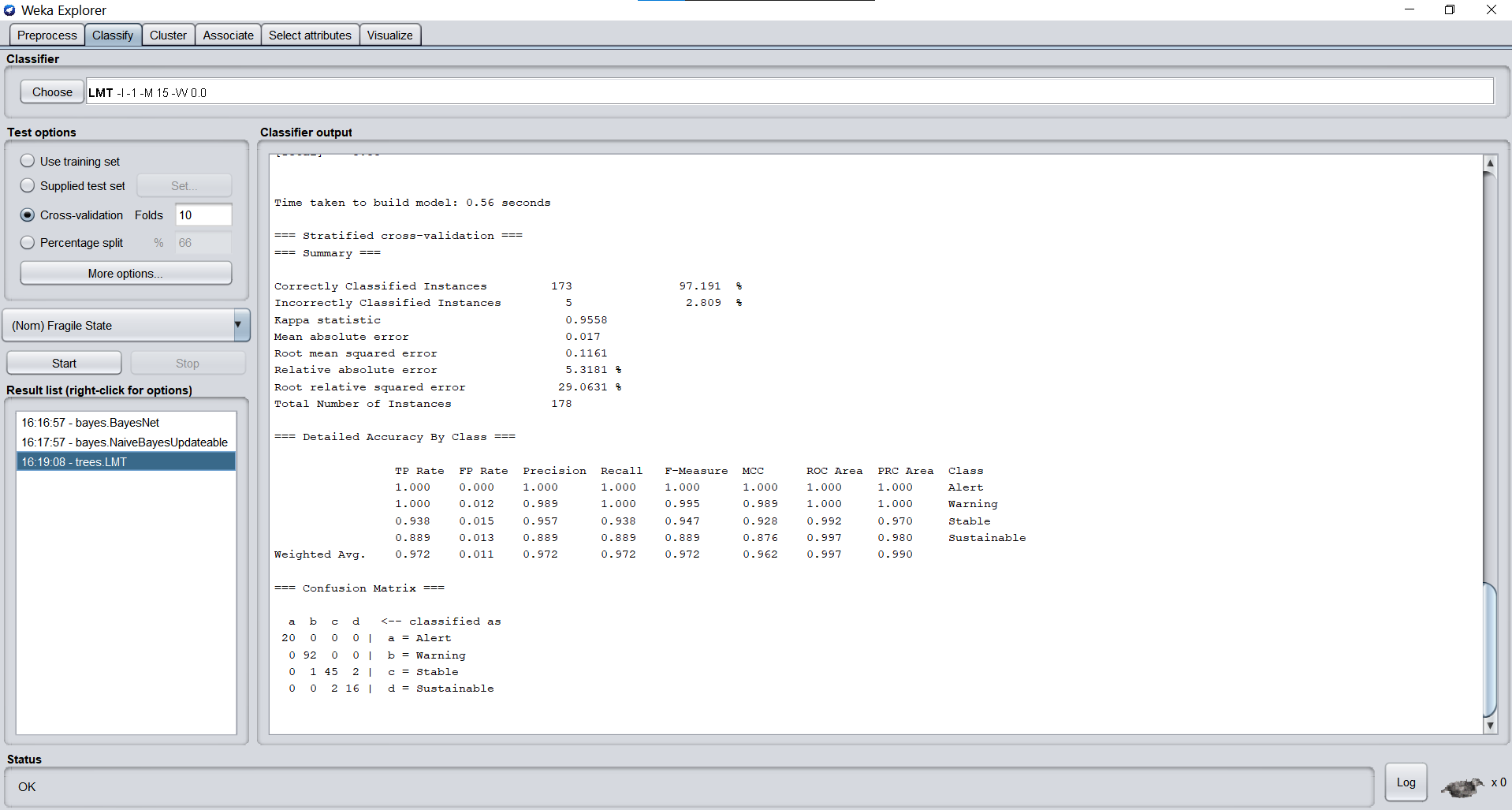
a b c d <-- classified as

20 0 0 0 | a = Alert

0 92 0 0 | b = Warning

0 1 45 2 | c = Stable

0 0 2 16 | d = Sustainable



**ACTION RULE MINING USING LISP MINER**

The software used to mine the action rules for a few selected attributes is known as the Lisp Miner. We have used two main types of attributes:

* **Stable Attributes**: These are the attributes whose value does not change i.e. fixed values.
* **Flexible Attributes**: These are the attributes that are different for each entry and have a possibity of changing.

There is also another attribute known as the Decision attribute.

The stable attributes that we have considered for this project are:

* Rank
* Country
* Year

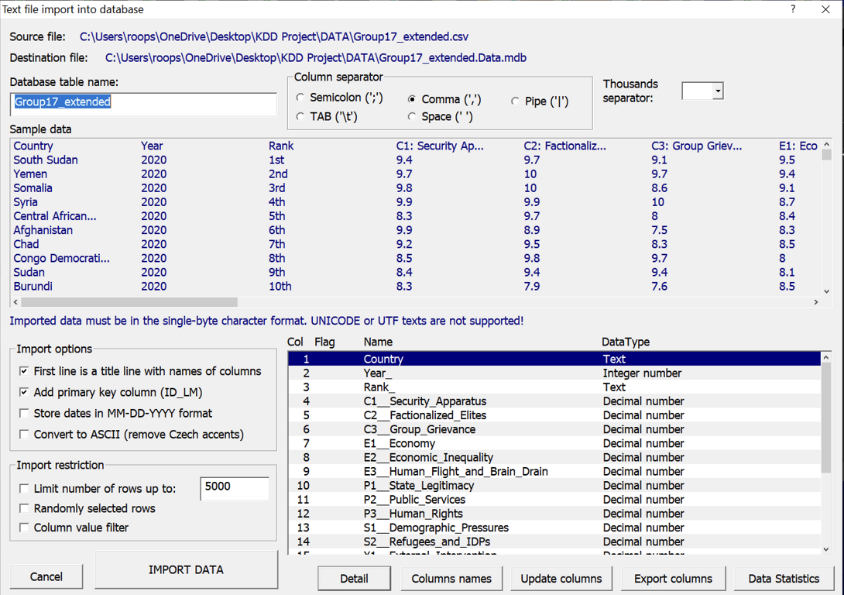
The Flexible attributes that we have considered for the mining are:

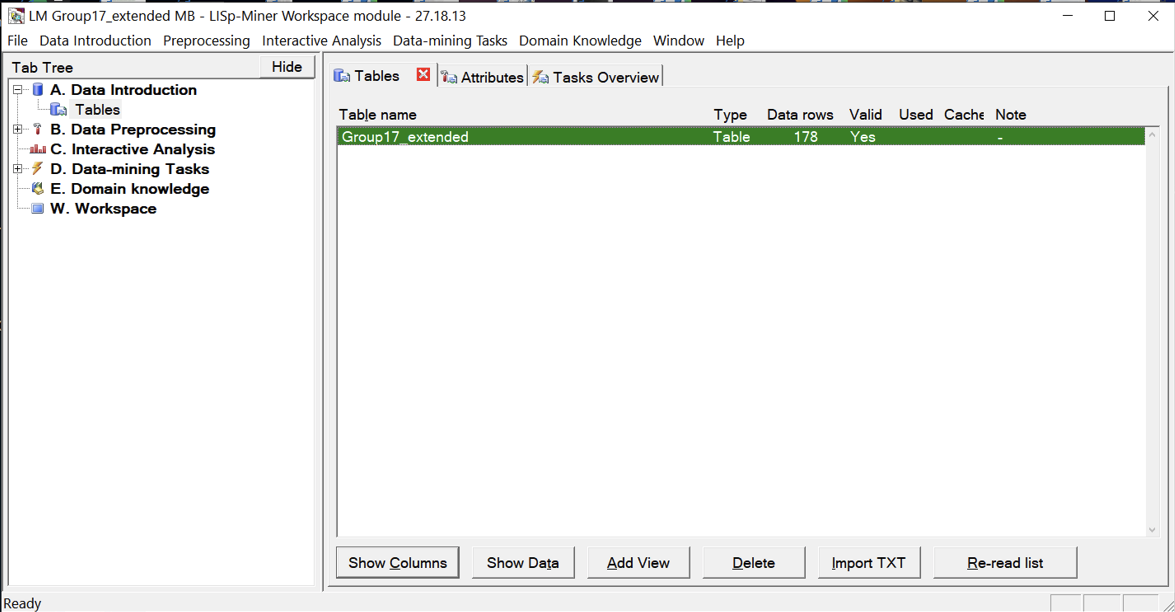
* C1: Security Apparatus
* Terrorism index
* Urban Population

The decision attribute of the project is Fragile State. The coefficient type is One Category and the type is from Alert to Stable.

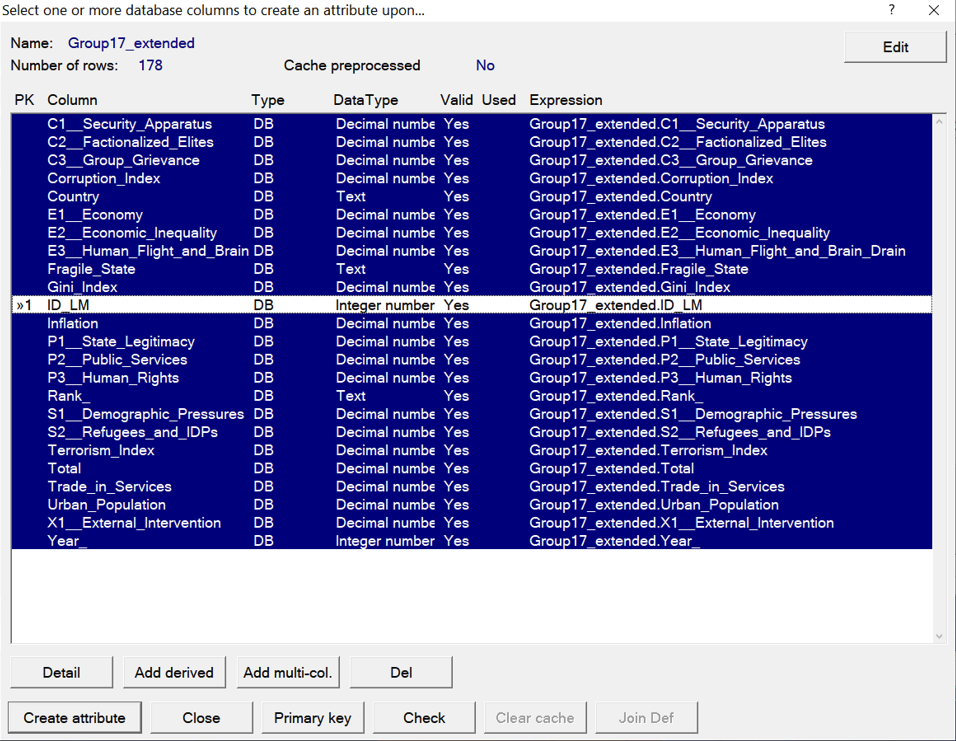
**STEPS FOLLOWED TO ACHIEVE RESULTS ON LISP MINER**:

1. Download, Unzip and Open LispMiner Application.
2. Select a file from the local machine and import the data.

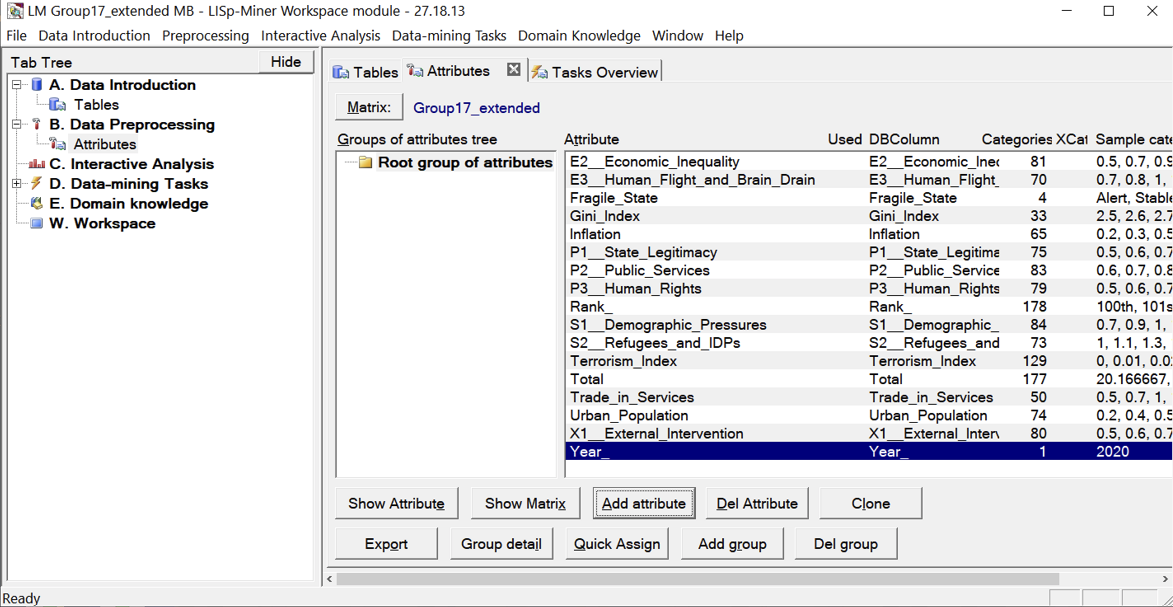


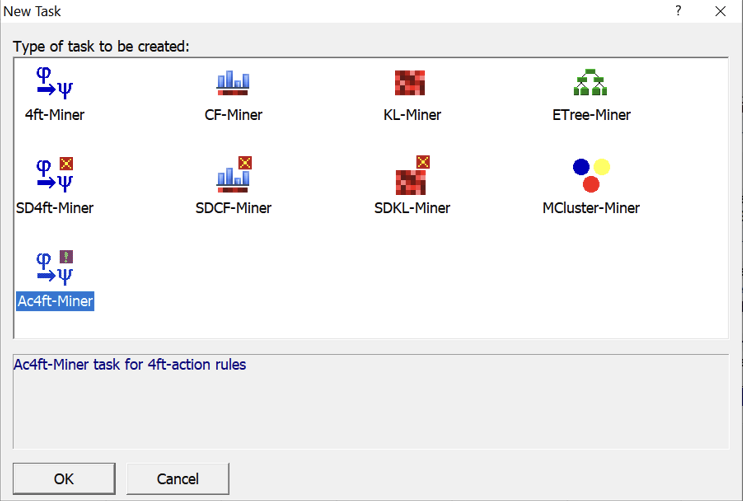


1. We have to select attributes. Hence, select all the attributes except the ID:



1. After build our attribues, right click on the Data Mining tasks and select create a new task. Now select the Ac4ft Miner.





1. Select the required stable and flexible attributes. Set the threshold values.

Graphical user interface, text, application

Description automatically generated

A screenshot of a computer

Description automatically generated

Press Run and wait for the results.

**RESULTS OF LISP MINER**

Graphical user interface, application, table, Excel

Description automatically generated

Graphical user interface, application, PowerPoint

Description automatically generated

Graphical user interface, application, PowerPoint

Description automatically generated

Graphical user interface, application, PowerPoint

Description automatically generated

**CONCLUSION**

The extended features that we have added to the original dataset have proved to be a bit efficient in the process of classification using the 3 main classifiers. So, there features can also be considered for the purposes of alerting the nations about their state. The action rules that were given out by the Lisp Miner that are required for a country that is in alert position to jump to a stable position are given. These rules can be analyzed and can a few changes can be made to the system by referring to these action rules. There can be a better way to analyze all the rules at once so that it would be easy to refer.

**REFERENCES**

**1) www.data.worldbank.org**

**2)** **https://fragilestatesindex.org/**

**3)** **www**.**tradingeconomics.com**

**4)** **https://en.wikipedia.org/wiki/Fragile\_States\_Index**