

Analysis on Fragile States Data

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

Instructor
Dr. Zbigniew W. Ras

TEAM MEMBERS:

Shreeya Gupta (801136226)
Yash Kanodia (801136232)
Anushka Tibrewal (801134231)
Ridhi Shah (801151577)

Problem Statement:

The problem statement describes the Fragile State Index as the lower the index score the better. To get a lower Index score, we need to discretize our Decision feature(i.e 'Total' column), which will then help us in finding the action rules. The action rules with the best support will give us the changes required in the classification feature to lower the index score, we are also required to add six new features in correlation with the data given. We will then analyze the results with and without adding the new features by running different classification algorithms on the data.

Introduction

In this project we were given Fragile State Index (FSI) dataset for all the countries which is available at <https://fragilestatesindex.org/excel/> and we have extended the dataset by adding 6 new features namely Inflation rate, Birth rate, Death Rate, GDP Rank, Life Expectancy, Economic Growth. We extracted this data from different sources.

Using the values for these newly added features, and using the feature Total as our decision feature with discretization we replace numbers with concepts such as:

1. Alert
2. Warning
3. Stable
4. Sustainable

Using the above concepts we find the best classifier for our dataset with the help of software WEKA and using classification algorithms like Random Forest, Simple Logistic and K-star. After finding the best classifier, we find the action rules and from these rules we can get the action which needs to be performed to change the state of the country from Alert to Stable or Sustainable.

Dataset

The dataset for the project was obtained from <https://fragilestatesindex.org/excel/> , we used the data for the year 2019. We downloaded the csv files of the features which correlated with our data from different source on the web and entered the new features to our dataset. The 6 new features are as follows:

1. Birth Rate : The ratio of live births in an area to the population of that area, expressed per 1000 population per area.
- 2.GDP Rank: It is the monetary market value of all final goods and services made within a country during a specific period.
3. Economic Growth : The economic growth for a country of one year, comparing one quarter of the country's gross domestic product to the previous quarter.
4. Death Rate: The ratio of live deaths in an area to the population of that area, expressed per 1000 population per area

5. Inflation Rate: The increase in consumer prices for the country in a year.
6. Life Expectancy : The average number of years to be lived by a group of people born in the same year, if mortality at each age remains constant in the future. Life expectancy at birth is also a measure of overall quality of life in a country and summarizes the mortality at all ages.

Motivation to select these features :

We searched for various features available on the Wikipedia, worldbank, and some other websites which provide us with different features which can fit with the previous features available in the Fragile State index 2019.

We performed different operations on six features, which we selected randomly and tried applying different classifiers to these features to check whether the accuracy increases or decreases based on the features selected.

After performing these operations, we got these six features:

1. Birth Rate
2. Death Rate
3. Gross Domestic Product (GDP)
4. Economic Growth
5. Inflation Rate
6. Life Expectancy

Adding these features to the previous data and applying different classifiers like random forest, Decision tree, simple Logistic Regression, Naïve Bayes, JRip and J48 on the data helped to improve the accuracy i.e the percentage of correctly classified class increases by a good margin and the percentage of incorrectly classified class decreases. After noting these accuracies, we performed operations to extract classification rules and action rules in this project. The extraction of classification and action rules are represented below.

DATA DISCRETIZATION (Using WEKA):

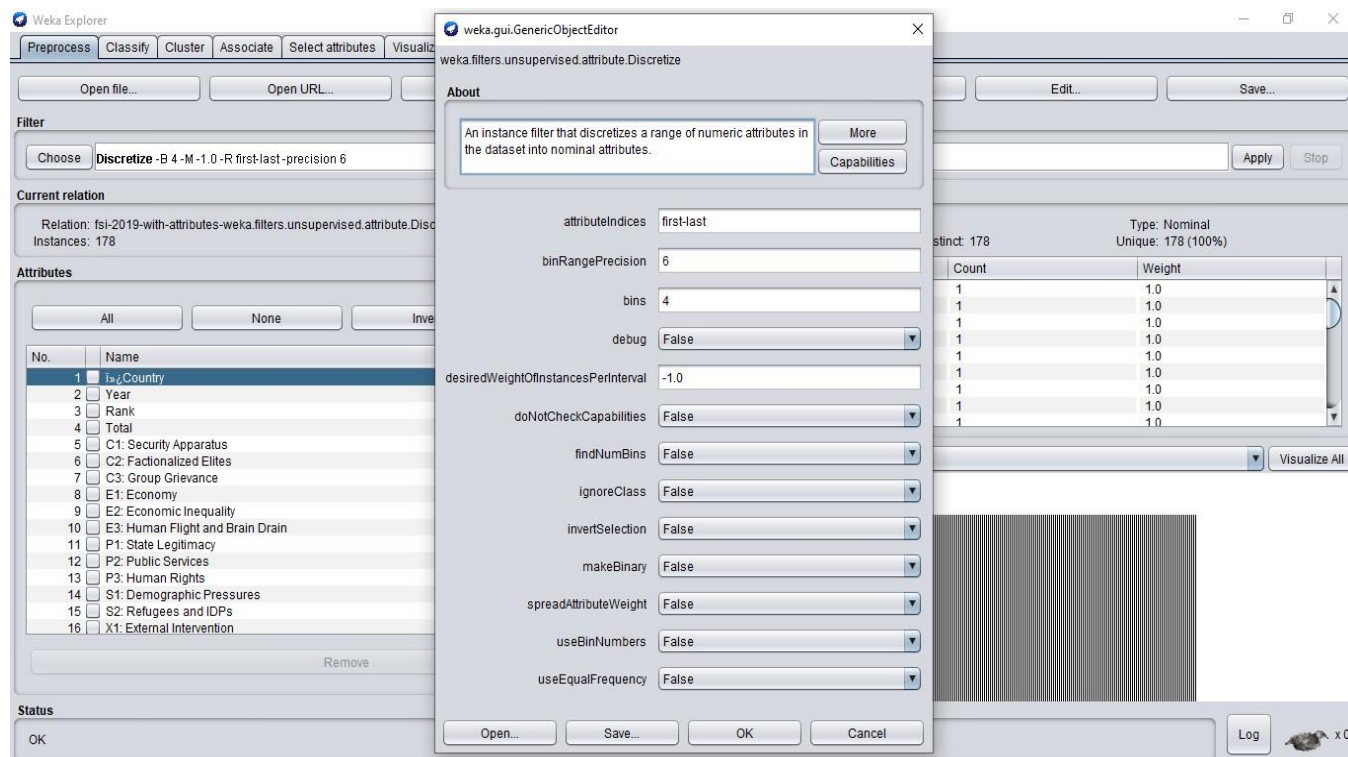
We used the WEKA tool for discretization as well as classification of our data.

Discretization alludes to the way toward changing over or parceling nonstop properties, highlights, or factors to discrete or ostensible qualities. 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.

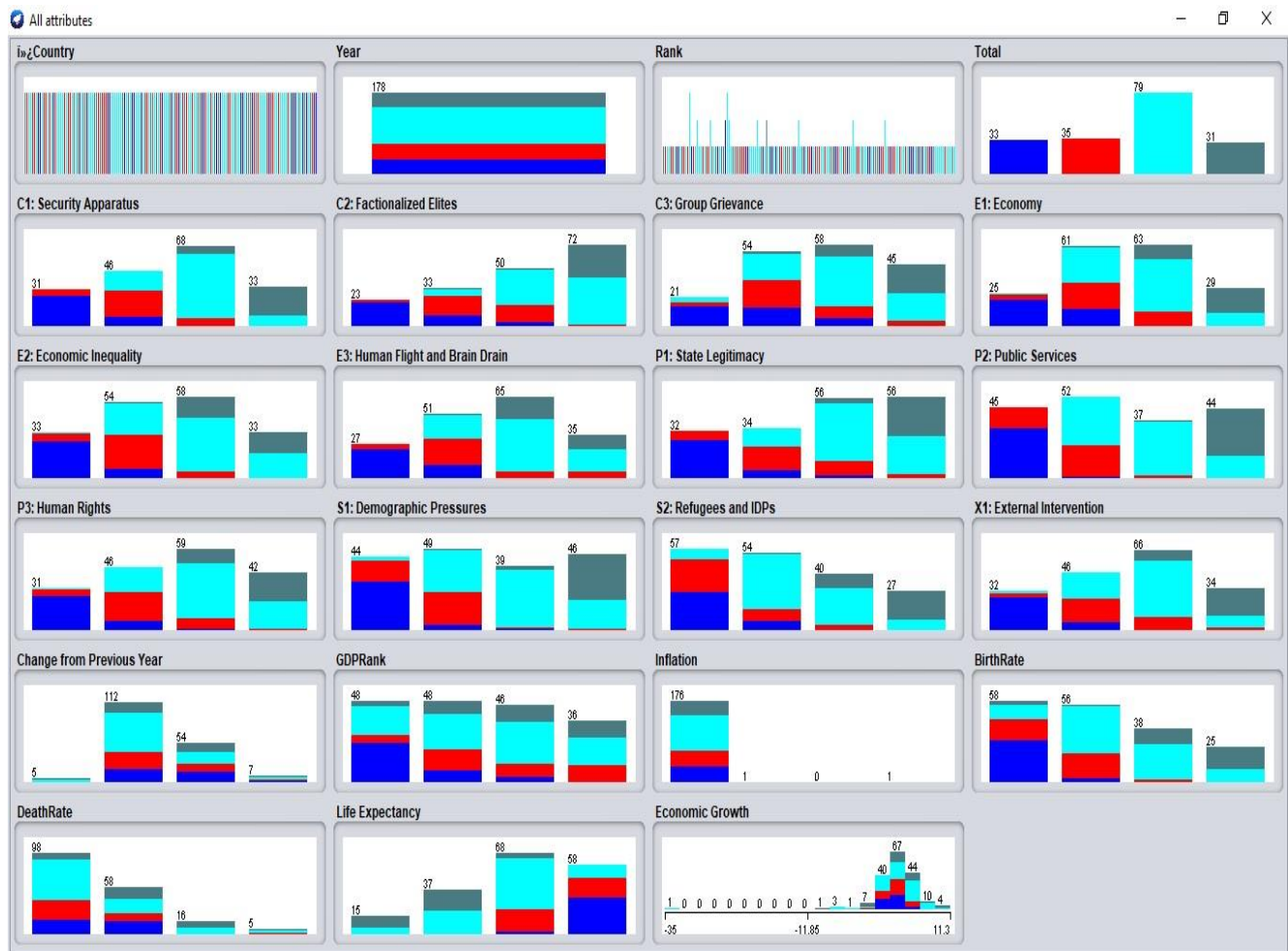
1. Alert: 90 - 120
2. Warning: 60 - 90
3. Stable: 30 - 60
4. Sustainable: 0 - 30

Source: https://en.wikipedia.org/wiki/Fragile_States_Index

We also had to discretize other attributes from numeric values to nominal values to perform effective classification.



Discretized data



DATA CLASSIFICATION (Using WEKA)

Data Classification is the way toward sorting out information by applicable classes with the goal that it might be utilized and ensured all the more proficiently. Following classification algorithms are used for data classification:

1. Naive Bayes: Naive Bayes is a classification algorithm. Traditionally it assumes that the input values are nominal, although it numerical inputs are supported by assuming a distribution. It is a basic strategy for building classifiers models that allocate class names to issue occasions, spoke to as vectors of highlight esteems, where the class marks are drawn from some limited set.
2. Random Forest: Random forests is the learning strategy for arrangement, relapse and different errands that works by building a large number of choice trees at preparing time and yielding the class that is the method of the classes (order) or mean expectation (regression) of the individual trees.
3. Decision Table: Decision Table classifier calculation is utilized to sum up the dataset by utilizing a decision table containing a similar number of properties as that of the first dataset. It is a short visual portrayal for indicating which activities to perform contingent upon given conditions.
4. JRip: It depends on affiliation rules with diminished error pruning (REP), extremely normal and powerful system found in decision tree calculations. At each phase of improvement, the pruning administrator picked is the one that yields the best decrease of mistake on the pruning set.

Class JRip java.lang.Object

weka.classifiers.AbstractClassifier weka.classifiers.rules.JRip

5. J48: The C4.5 calculation for building decision trees is executed in Weka as a classifier called J48. Envision that you have a dataset with a rundown of indicators or free factors and a rundown of targets or ward factors. Decision tree J48 is the usage of calculation Iterative Dichotomiser 3.

1) Naive Bayes:

Without adding new features-

The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```

Correctly Classified Instances      165      92.6966 %
Incorrectly Classified Instances    13      7.3034 %
Kappa statistic                    0.8954
Mean absolute error                0.054
Root mean squared error            0.1731
Relative absolute error            15.41 %
Root relative squared error        41.3866 %
Total Number of Instances         178
Ignored Class Unknown Instances    1921

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.970  0.034  0.865  0.970  0.914  0.896  0.980  0.954  '(-inf-41.05]'
0.800  0.014  0.933  0.800  0.862  0.835  0.919  0.894  '(41.05-65.2]'
0.962  0.040  0.950  0.962  0.956  0.921  0.976  0.970  '(65.2-89.35]'
0.935  0.014  0.935  0.935  0.935  0.922  0.970  0.961  '(89.35-inf]'
Weighted Avg.  0.927  0.029  0.928  0.927  0.926  0.899  0.964  0.951

=== Confusion Matrix ===

 a b c d <-- classified as
32 1 0 0 | a = '(-inf-41.05]'
 5 28 2 0 | b = '(41.05-65.2]'
 0 1 76 2 | c = '(65.2-89.35]'

```

The 'Result list' on the left shows a single entry: '19:57:59 - bayes.NaiveBayes'.

With adding new features:

The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```

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

Correctly Classified Instances      160      89.8876 %
Incorrectly Classified Instances    18     10.1124 %
Kappa statistic                    0.8568
Mean absolute error                0.0604
Root mean squared error            0.1991
Relative absolute error            17.2284 %
Root relative squared error        47.5952 %
Total Number of Instances         178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0.939  0.028  0.886  0.939  0.912  0.892  0.954  0.974  '(-inf-41.05]'
0.829  0.014  0.935  0.829  0.879  0.854  0.975  0.938  '(41.05-65.2]'
0.899  0.040  0.947  0.899  0.922  0.864  0.979  0.972  '(65.2-89.35]'
0.935  0.054  0.784  0.935  0.853  0.823  0.967  0.926  '(89.35-inf]'
Weighted Avg.  0.899  0.035  0.905  0.899  0.900  0.860  0.963  0.958

=== Confusion Matrix ===

 a b c d <-- classified as
31 2 0 0 | a = '(-inf-41.05]'
 4 29 2 0 | b = '(41.05-65.2]'
 0 0 71 8 | c = '(65.2-89.35]'
 0 0 2 29 | d = '(89.35-inf]'

```

The 'Result list' on the left shows three entries: '19:19:47 - bayes.NaiveBayes', '19:28:38 - bayes.NaiveBayes', and '19:30:22 - bayes.NaiveBayes'.

2) Random Forest:

Without adding new features:

Weka Explorer interface showing the Random Forest classifier results. The classifier is configured with parameters: -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1. The test options are set to Cross-validation with 10 folds. The classifier output shows the following summary statistics:

Metric	Value	Percentage
Correctly Classified Instances	128	71.9101 %
Incorrectly Classified Instances	50	28.0899 %
Kappa statistic	0.5599	
Mean absolute error	0.2675	
Root mean squared error	0.3295	
Relative absolute error	76.2989 %	
Root relative squared error	78.7665 %	
Total Number of Instances	178	
Ignored Class Unknown Instances	1921	

The detailed accuracy by class is as follows:

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
'(-inf-41.05]'	0.848	0.034	0.848	0.848	0.848	0.814	0.999	0.955
'(41.05-65.2]'	0.086	0.014	0.600	0.086	0.150	0.173	0.956	0.743
'(65.2-89.35]'	1.000	0.434	0.648	1.000	0.786	0.605	0.997	0.957
'(89.35-inf)'	0.581	0.000	1.000	0.581	0.735	0.730	0.999	0.952
Weighted Avg.	0.719	0.202	0.737	0.719	0.664	0.581	0.990	0.914

The confusion matrix is as follows:

```

=== Confusion Matrix ===
 a b c d <-- classified as
28 2 3 0 | a = '(-inf-41.05]
 5 3 27 0 | b = '(41.05-65.2]
 0 0 79 0 | c = '(65.2-89.35]

```

With adding new features:

Weka Explorer interface showing the Random Forest classifier results with new features. The classifier is configured with parameters: -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1. The test options are set to Cross-validation with 10 folds. The classifier output shows the following summary statistics:

Metric	Value	Percentage
Correctly Classified Instances	139	78.0899 %
Incorrectly Classified Instances	39	21.9101 %
Kappa statistic	0.665	
Mean absolute error	0.2459	
Root mean squared error	0.3077	
Relative absolute error	70.155 %	
Root relative squared error	73.5532 %	
Total Number of Instances	178	

The detailed accuracy by class is as follows:

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
'(-inf-41.05]'	0.939	0.034	0.861	0.939	0.899	0.875	0.991	0.961
'(41.05-65.2]'	0.257	0.014	0.818	0.257	0.391	0.401	0.965	0.869
'(65.2-89.35]'	1.000	0.323	0.712	1.000	0.832	0.694	0.982	0.977
'(89.35-inf)'	0.645	0.000	1.000	0.645	0.784	0.775	0.991	0.964
Weighted Avg.	0.781	0.153	0.811	0.781	0.749	0.684	0.982	0.950

The confusion matrix is as follows:

```

=== Confusion Matrix ===
 a b c d <-- classified as
31 2 0 0 | a = '(-inf-41.05]
 5 9 21 0 | b = '(41.05-65.2]
 0 0 79 0 | c = '(65.2-89.35]
 0 0 11 20 | d = '(89.35-inf)

```

3) Decision Table:

Without adding new features:

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose: **DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"**

Test options

☐ Use training set
☐ Supplied test set (Set...)
☒ Cross-validation Folds: **10**
☐ Percentage split %: **66**
 More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 19:57:59 - bayes.NaiveBayes
- 20:01:14 - trees.J48
- 20:03:31 - trees.RandomForest
- 20:05:32 - rules.DecisionTable**

Classifier output

```

Correctly Classified Instances      134      75.2809 %
Incorrectly Classified Instances    44      24.7191 %
Kappa statistic                    0.6455
Mean absolute error                0.2014
Root mean squared error            0.2986
Relative absolute error            57.454 %
Root relative squared error        71.3624 %
Total Number of Instances         178
Ignored Class Unknown Instances    1921

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.879    0.083    0.707     0.879    0.784     0.735    0.068     0.009    '(-inf-41.05]'
          0.457    0.063    0.640     0.457    0.533     0.451    0.986     0.588    '(41.05-65.2]'
          0.835    0.152    0.815     0.835    0.825     0.682    0.995     0.875    '(65.2-89.35]'
          0.742    0.054    0.742     0.742    0.742     0.688    0.982     0.770    '(89.35-inf)'
Weighted Avg.  0.753    0.104    0.748     0.753    0.746     0.647    0.819     0.640

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
29  4  0  0 | a = '(-inf-41.05]'
11 16  8  0 | b = '(41.05-65.2]'
 0  5 66  8 | c = '(65.2-89.35]'
  
```

Status: OK Log x 0

With adding new features-

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose: **DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"**

Test options

☐ Use training set
☐ Supplied test set (Set...)
☒ Cross-validation Folds: **10**
☐ Percentage split %: **66**
 More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 19:50:50 - trees.RandomForest
- 19:54:03 - rules.DecisionTable**

Classifier output

```

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

Correctly Classified Instances      120      67.4157 %
Incorrectly Classified Instances    58      32.5843 %
Kappa statistic                    0.5348
Mean absolute error                0.2464
Root mean squared error            0.331
Relative absolute error            70.278 %
Root relative squared error        79.1129 %
Total Number of Instances         178

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.879    0.124    0.617     0.879    0.725     0.665    0.952     0.882    '(-inf-41.05]'
          0.343    0.070    0.545     0.343    0.421     0.330    0.812     0.512    '(41.05-65.2]'
          0.734    0.212    0.734     0.734    0.734     0.522    0.844     0.771    '(65.2-89.35]'
          0.677    0.061    0.700     0.677    0.689     0.624    0.945     0.772    '(89.35-inf)'
Weighted Avg.  0.674    0.142    0.669     0.674    0.663     0.529    0.875     0.741

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
29  4  0  0 | a = '(-inf-41.05]'
12 12 11  0 | b = '(41.05-65.2]'
 6  6 58  9 | c = '(65.2-89.35]'
 0  0 10 21 | d = '(89.35-inf)'
  
```

Status: OK Log x 0

4)JRip:

Without adding new features:

The screenshot shows the Weka Explorer interface with the JRip classifier selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Classifier output

Correctly Classified Instances	126	70.7865 %
Incorrectly Classified Instances	52	29.2135 %
Kappa statistic	0.5743	
Mean absolute error	0.1686	
Root mean squared error	0.3476	
Relative absolute error	48.105 %	
Root relative squared error	83.0867 %	
Total Number of Instances	178	
Ignored Class Unknown Instances	1921	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.697	0.055	0.742	0.697	0.719	0.658	0.931	0.683	'(-inf-41.05]'
	0.543	0.105	0.559	0.543	0.551	0.443	0.829	0.407	'(41.05-65.2]'
	0.810	0.242	0.727	0.810	0.766	0.564	0.432	0.033	'(65.2-89.35]'
	0.645	0.034	0.800	0.645	0.714	0.667	0.860	0.603	'(89.35-inf)'
Weighted Avg.	0.708	0.144	0.710	0.708	0.706	0.576	0.677	0.326	

==== Confusion Matrix ====

```

a b c d <-- classified as
23 7 3 0 | a = '(-inf-41.05]
6 19 10 0 | b = '(41.05-65.2]
2 8 64 5 | c = '(65.2-89.35]

```

The 'Result list' on the left shows several models, with '20:07:15 - rules.JRip' selected.

With adding new features:

The screenshot shows the Weka Explorer interface with the JRip classifier selected. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Classifier output

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

==== Summary ====

Correctly Classified Instances	136	76.4045 %
Incorrectly Classified Instances	42	23.5955 %
Kappa statistic	0.6578	
Mean absolute error	0.1418	
Root mean squared error	0.323	
Relative absolute error	40.4439 %	
Root relative squared error	77.2038 %	
Total Number of Instances	178	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.848	0.055	0.778	0.848	0.812	0.768	0.910	0.718	'(-inf-41.05]'
	0.543	0.070	0.655	0.543	0.594	0.509	0.762	0.512	'(41.05-65.2]'
	0.848	0.192	0.779	0.848	0.812	0.652	0.861	0.759	'(65.2-89.35]'
	0.710	0.034	0.815	0.710	0.759	0.714	0.905	0.698	'(89.35-inf)'
Weighted Avg.	0.764	0.115	0.761	0.764	0.760	0.656	0.858	0.692	

==== Confusion Matrix ====

```

a b c d <-- classified as
28 5 0 0 | a = '(-inf-41.05]
6 19 10 0 | b = '(41.05-65.2]
2 5 67 5 | c = '(65.2-89.35]
0 0 9 22 | d = '(89.35-inf)

```

The 'Result list' on the left shows several models, with '19:56:30 - rules.JRip' selected.

5)J48:

Without adding new features:

Classifier

Choose **J48 - C 0.25 - M 2**

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66
 More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 19:57:59 - bayes.NaiveBayes
- 20:01:14 - trees.J48

Classifier output

```

Correctly Classified Instances      139          78.0899 %
Incorrectly Classified Instances    39          21.9101 %
Kappa statistic                    0.6854
Mean absolute error                 0.1201
Root mean squared error             0.3062
Relative absolute error             34.2598 %
Root relative squared error         73.1816 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
      0.788    0.055    0.765    0.788    0.776    0.724    0.794    0.703    '(-inf-41.05]'
      0.686    0.084    0.667    0.686    0.676    0.595    0.777    0.462    '(41.05-65.2]'
      0.861    0.131    0.840    0.861    0.850    0.728    0.862    0.749    '(65.2-89.35]'
      0.677    0.041    0.778    0.677    0.724    0.673    0.842    0.683    '(89.35-inf]'
Weighted Avg.    0.781    0.092    0.781    0.781    0.780    0.692    0.829    0.672

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
26  7  0  0 | a = '(-inf-41.05]'
 8 24  3  0 | b = '(41.05-65.2]'
 0  5 68  6 | c = '(65.2-89.35]'

```

Status

OK Log x 0

With adding new features:

Classifier

Choose **J48 - C 0.25 - M 2**

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66
 More options...

(Nom) Total

Start Stop

Result list (right-click for options)

- 19:50:50 - trees.RandomForest
- 19:54:03 - rules.DecisionTable
- 19:56:30 - rules.JRip
- 20:00:15 - trees.J48

Classifier output

```

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

Correctly Classified Instances      138          77.5281 %
Incorrectly Classified Instances    40          22.4719 %
Kappa statistic                    0.6808
Mean absolute error                 0.1248
Root mean squared error             0.3079
Relative absolute error             35.6156 %
Root relative squared error         73.5851 %
Total Number of Instances          178

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
      0.697    0.062    0.719    0.697    0.708    0.643    0.883    0.727    '(-inf-41.05]'
      0.629    0.105    0.595    0.629    0.611    0.513    0.859    0.569    '(41.05-65.2]'
      0.848    0.091    0.882    0.848    0.865    0.761    0.908    0.851    '(65.2-89.35]'
      0.839    0.048    0.788    0.839    0.813    0.772    0.896    0.700    '(89.35-inf]'
Weighted Avg.    0.775    0.081    0.779    0.775    0.777    0.692    0.892    0.746

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
23 10  0  0 | a = '(-inf-41.05]'
 9 22  4  0 | b = '(41.05-65.2]'
 0  5 67  7 | c = '(65.2-89.35]'
 0  0  5 26 | d = '(89.35-inf]'

```

Status

OK Log x 0

RESULTS

Classifier Used on the data	Before Adding Attributes		After Adding Attributes	
	Correctly Classified	Incorrectly Classified	Correctly Classified	Incorrectly Classified
Bayes Net	92%	8%	91%	9%
Decision Table	75%	25%	67.5%	32.5%
J48	78%	22%	77.5%	22.5%
JRip	71%	29%	76.4%	23.6%
Naïve Bayes	93%	7%	89.9%	10.1%
Random Forest	72%	28%	78%	22%

When these classifiers are applied on the data (without new features), the accuracy which is represented by the correctly classified and incorrectly classified classes was around 85% and when the new features were added in the data the accuracy got decrease by 4-5% in 4 classifiers and in 2 classifiers it got improved.

EXTRACTING ACTION RULES USING LISP MINER

Introduction

The LISP-Miner system is an academic data mining software tool developed at the University of Economics, Prague¹, which is focused on mining various types of association rules from categorical data. LISP-Miner offers a greater variety of different types of relations between the left-hand and right-hand sides of a rule using various data mining procedures implemented by LISP-miner. For the purpose of this project, we are using Ac4ft-Miner data mining procedure for extracting action rules. Ac4ft-Miner finds rules that express which actions should be performed to improve the defined state. It achieves it by examining the dependencies among the data given as an input.

The process of action rule discovery used by Lisp Miner:

LISP-Miner implements various GUHA procedures that mine for different types of knowledge patterns. This system consists of 10 different data mining procedures where 4ft-Miner procedure is based on the original GUHA procedure ASSOC while other procedures have been designed during the development of the system.

The 4ft-Miner procedure mines for knowledge patterns, that can be understood as 4ft association rules of the form

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

where ϕ (antecedent), ψ (succedent) and γ (condition) are cedents and \approx is a quantifier which is evaluated on the subset of examples, that satisfy the condition γ .

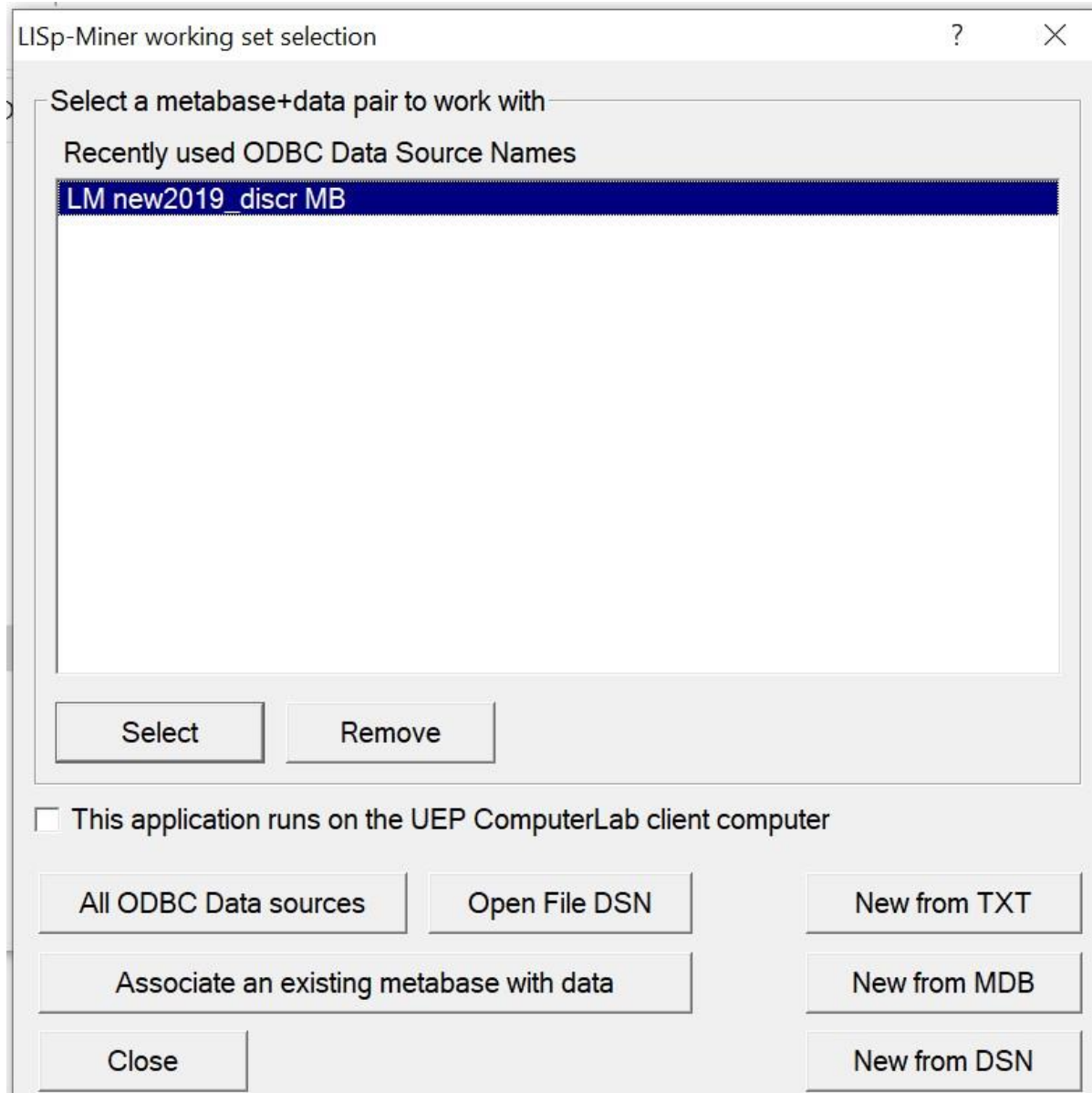
The work with the LISP-Miner is not so straightforward as with other data mining systems because LISPMiner is distributed as a number of executables that must be invoked by the user. Each mining procedure in LISP-Miner tool uses different processing modules as:

1. Module 1 - LMAdmin: The first module to be used is the LMAdmin. The primary aim of this module is to attach the meta-data base to the analyzed data. The concept of meta-data allows to store and reuse inputs for the tasks as well the results obtained during analysis. Both data and meta-data are stored using a database. This is a necessary step that must precede any analysis.
2. Module 2 – LMDataSource: This module implements a variety of data transformation and data preprocessing methods that can be used to select attributes for the specific data mining task, create derived attributes, or discretize numeric attributes.
3. Module 3 - Data processing module (Task): This is task module which is used to analyze the data, and creates the task using respective xxxTask module.
4. Module 4 - Data interpretation module (Result): This is Result module which is used to display and evaluate the results. This is used to display, sort or select the resulting rules generated by running the corresponding xxxResult module.

Steps to extract action rules using LISP miner :

Step 1 : First step is to acquire dataset and manipulate the data by adding 6 new features.

Step 2 : Import dataset into LISP miner tool using 'New from TXT' tab. For importing data through text file, file should be of type text or csv.



Step 3 : After uploading dataset to miner tool, import data from CSV file into database. At this stage, LISP miner displays all the attribute names as mentioned in dataset file.

Text file import into database

Source file: D:\ABIA\Exam\new2019_discr.csv
Destination file: D:\ABIA\Exam\new2019_discr.Data.mdb

Database table name: new2019_discr

Column separator: ☐ Semicolon (';') ☒ Comma (',') ☐ Pipe ('|')
☐ TAB ('\t') ☐ Space (' ') Thousands separator:

Sample data

	Country	Year	Rank	Total	'C1: Security A...	'C2: Fa
0	Afghanistan	'\AI\'	9th	Alert	'\{7.675-inf)\...	'\{7.75
1	Albania	'\AI\'	121st	Stable	'\{3.025-5.35]...	'\{5.5-
2	Algeria	'\AI\'	72nd	Warning	'\{5.35-7.675]...	'\{5.5-
3	Angola	'\AI\'	35th	Warning	'\{5.35-7.675]...	'\{5.5-
4	'Antigua and Ba...	'\AI\'	127th	Stable	'\{5.35-7.675]...	'\{3.25
5	Argentina	'\AI\'	140th	Stable	'\{3.025-5.35]...	'\{-inf-
6	Armenia	'\AI\'	105th	Warning	'\{3.025-5.35]...	'\{5.5-
7	Australia	'\AI\'	174th	Sustainable	'\{-inf-3.025]...	'\{-inf-
8	Austria	'\AI\'	165th	Sustainable	'\{-inf-3.025]...	'\{-inf-

Imported data must be in the single-byte character format. UNICODE or UTF texts are not supported!

Import options

- ☒ First line is a title line with names of columns
- ☒ Add primary key column (ID_LM)
- ☐ Store dates in MM-DD-YYYY format
- ☐ Convert to ASCII (remove Czech accents)

Import restriction

- ☐ Limit number of rows up to: 5000
- ☐ Randomly selected rows
- ☐ Column value filter

Col Flag Name DataType

1		Field01	Integer number
2		Country	Text
3		Year	Text
4		Rank	Text
5		Total	Text
6	x	C1_Security_Apparatus	Text
7	x	C2_Factionalized_Elites	Text
8	x	C3_Group_Grievance	Text
9	x	E1_Economy	Text
10	x	E2_Economic_Inequality	Text
11	x	E3_Human_Flight_and_Brain_Drain	Text
12	x	P1_State_Legitimacy	Text
13	x	P2_Public_Services	Text
14	x	P3_Human_Rights	Text
15	x	C4_Democratic_Resources	Text

Cancel IMPORT DATA Detail Columns names Update columns Export columns Data Statistics

Step 4 : Select attributes from given list to extract rules from uploaded database. These attributes can be used to decide stable attributes, variable attributes and decision attributes.

Select Decision attribute: Selecting Total attribute in the Decision Group

LM new2019_discr MB - LISP-Miner Workspace module - 27.18.08

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

Matrix: new2019_discr

Groups of attributes tree

- Root group of attribute
- Decision
- Flexible
- Stable

Attribute basic parameters

Name: Total

Derived from the name of column ☒

Name for easy interpretation

Name in literals: Total

Derived from the name of attribute ☒

Show name in literals ☒

Nominal/Ordinal/Cardinal type: Nominal

Variability type: Not Set

Row shift in the analysed data matrix (>0 ... fwd; <0 ... bwd): 0

Comment: -

OK Cancel

Show Attribute Show Matrix Add attribute Del Attribute Clone

Export Group detail Quick Assign Add group Del group

Step 5: Select Flexible attribute: Selecting Security_Apparatus, Economy, Demographic_Pressure, Refugees_and_IDPs, External_Intervention attribute in the Flexible Group

The screenshot shows the LMSp-Miner Workspace module with the 'Attributes' tab selected. The 'Groups of attributes tree' on the left shows a hierarchy: Root group of attributes, Decision, Flexible, and Stable. The 'Flexible' group is selected, and the 'Attributes' tab is active. The main table displays the selected attributes and their properties.

Attribute	Used	DBCColumn	Categories	XCat	Sample categories
x_C1_Security_Apparatus	x_C1_Security_Ap	4	\(3.025-5.35)\", \((5.35-7.675)\", \((7.675-		
x_E1_Economy	x_E1_Economy	4	\(3.35-5.5)\", \((5.5-7.65)\", \((7.65-inf)\",		
x_S1_Demographic_Pressures	x_S1_Demographi	4	\(3.25-5.5)\", \((5.5-7.75)\", \((7.75-inf)\",		
x_S2_Refugees_and_IDPs	x_S2_Refugees ar	4	\(3.325-5.55)\", \((5.55-7.775)\", \((7.775-		
x_X1_External_Intervention	x_X1_External inte	4	\(3.025-5.35)\", \((5.35-7.675)\", \((7.675-		

Buttons at the bottom: Show Attribute, Show Matrix, Add attribute, Del Attribute, Clone, Export, Group detail, Quick Assign, Add group, Del group.

Step 6:

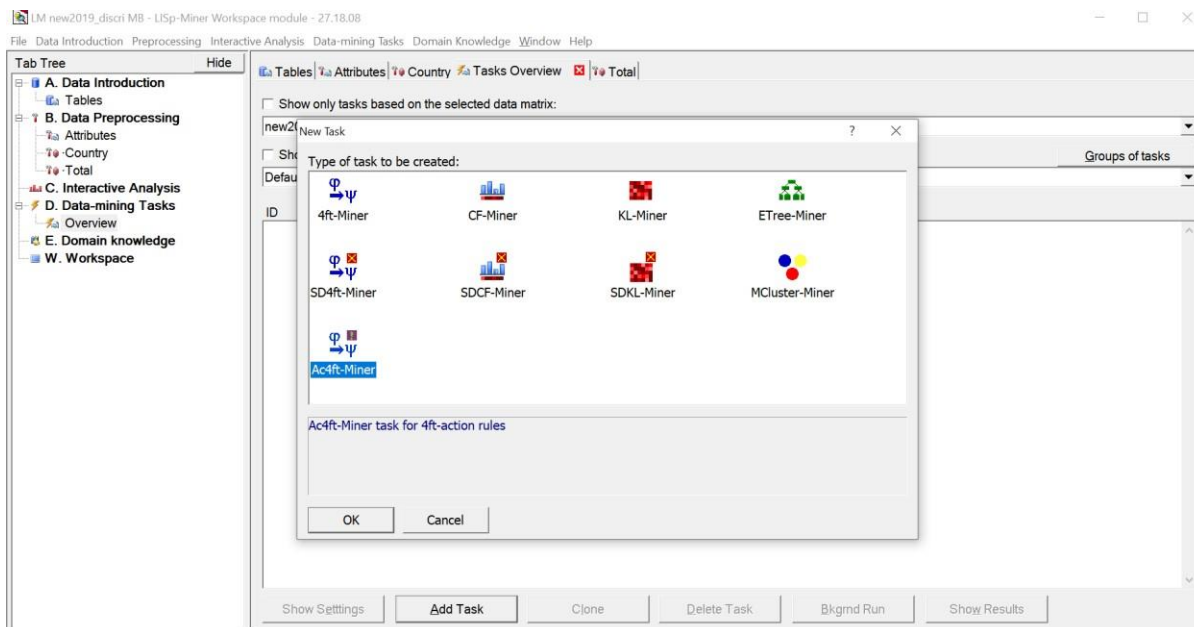
Selecting Stable attribute: Selecting Country attribute in Stable group

The screenshot shows the LMSp-Miner Workspace module with the 'Attributes' tab selected. The 'Groups of attributes tree' on the left shows a hierarchy: Root group of attributes, Decision, Flexible, and Stable. The 'Stable' group is selected, and the 'Attributes' tab is active. The main table displays the selected attributes and their properties.

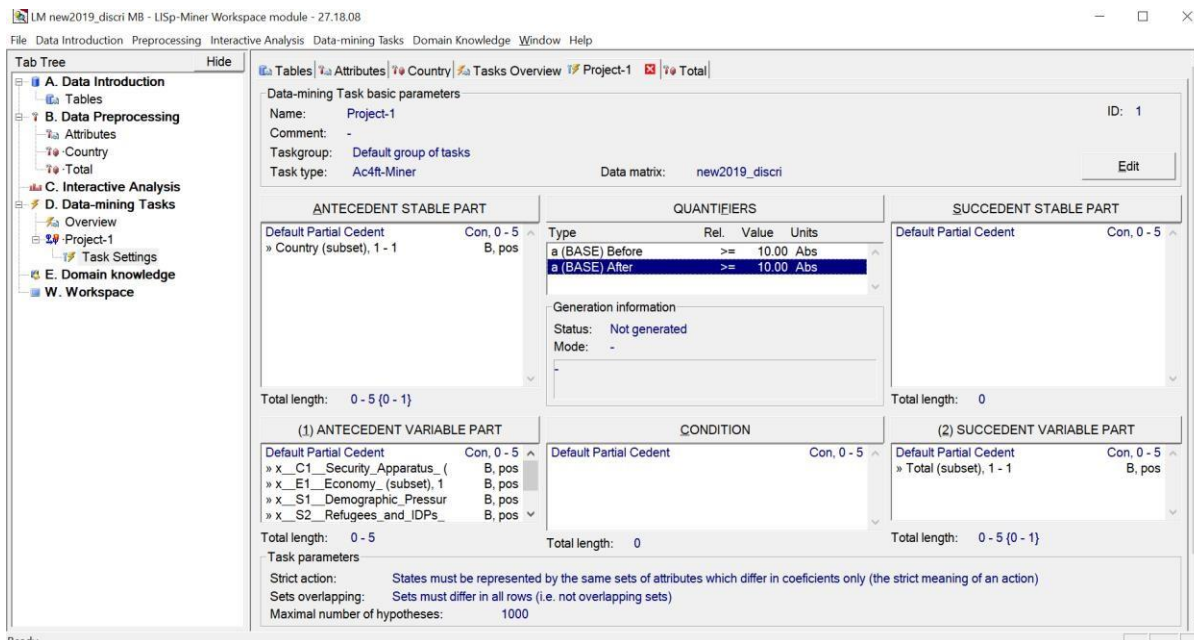
Attribute	Used	DBCColumn	Categories	XCat	Sample categories
Country	Country	177	Alghanistan, Albania, Algeria, Angola, 'Ar		

Buttons at the bottom: Show Attribute, Show Matrix, Add attribute, Del Attribute, Clone, Export, Group detail, Quick Assign, Add group, Del group.

Step 7 : Create new task using Ac4ft-Miner data mining procedure for extracting rules.



Step 8: Generate tasks: Selected Country attribute in the Antecedent stable part, Security_Apparatus, Economy, Demographic_Pressure, Refugees_and_IDPs, External_Intervention in the Antecedent variable part. Selected Total in the Succedent Variable part with the one category constraint having min value 1 and max value 2. A threshold of 10 in before and after value in the Quantifiers.



Step 9 : Once stable, variable & decision attributes, conditions, etc. are selected , run the query to generate rules. LISP miner tool displays list of all the rules generated based on attribute selection

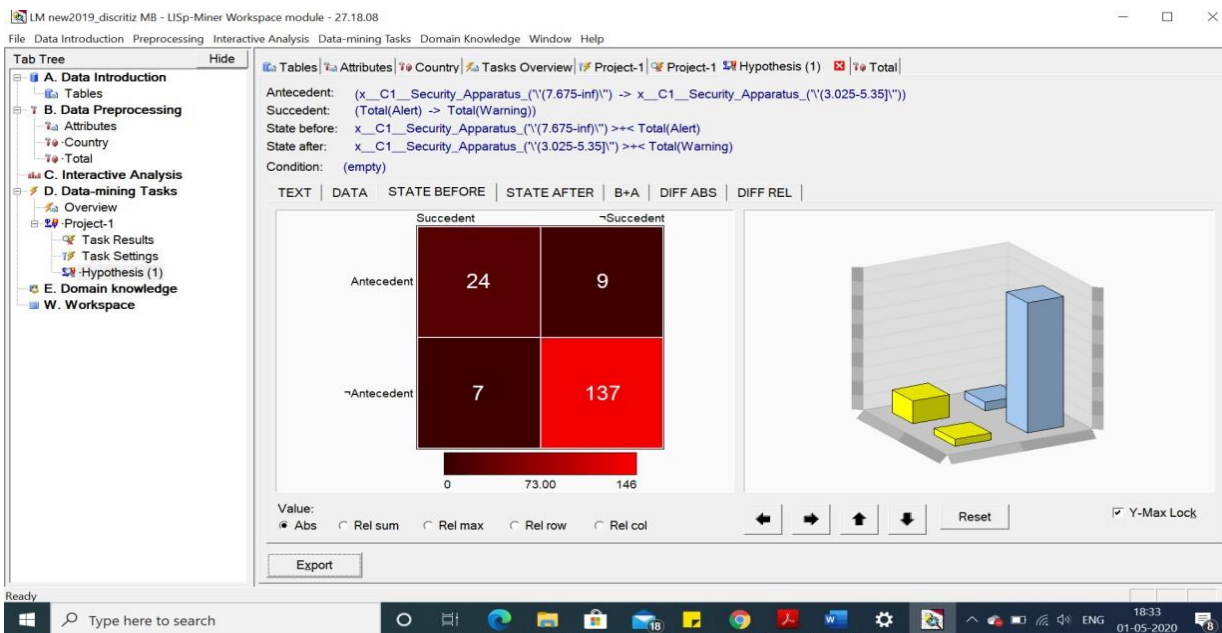
The screenshot shows the LISP-Miner Workspace module. The main workspace displays the following information:

- Task:** Project-1
- Comment:** -
- Taskgroup:** Default group of tasks
- Data matrix:** new2019_discr
- Task type:** Ac4R-Miner
- Task run:**
 - Start: 30.4.2020 19:09:47
 - Total time: 0h 0m 0s
 - Number of verifications: 256
 - Number of hypotheses: 29
 - Mode: Standard
- Actual group of hypotheses:** All hypotheses
- Hypotheses in group:** 29
- Shown hypotheses:** 29
- Highlighted:** 0

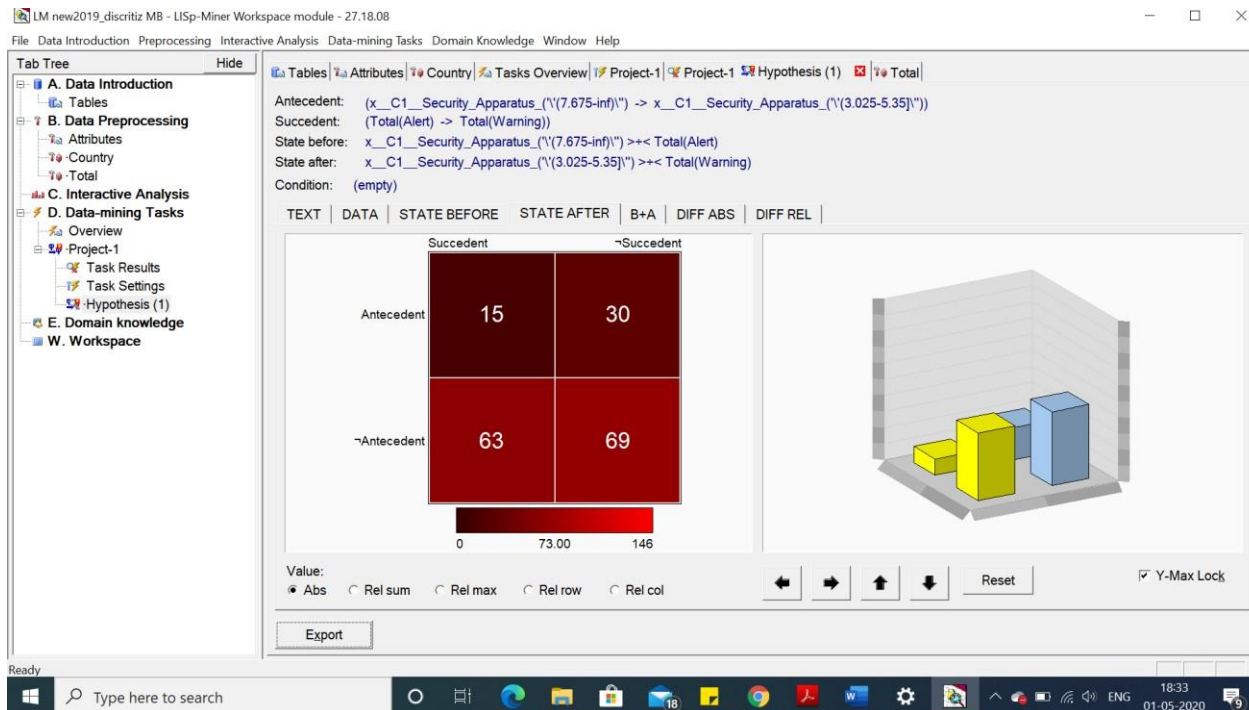
The list of hypotheses is displayed in a table with columns: Nr, Id, Df-Conf, B-Conf, A-Conf, and Hypothesis. The hypotheses are generated based on the selected attributes and conditions.

Visualization:

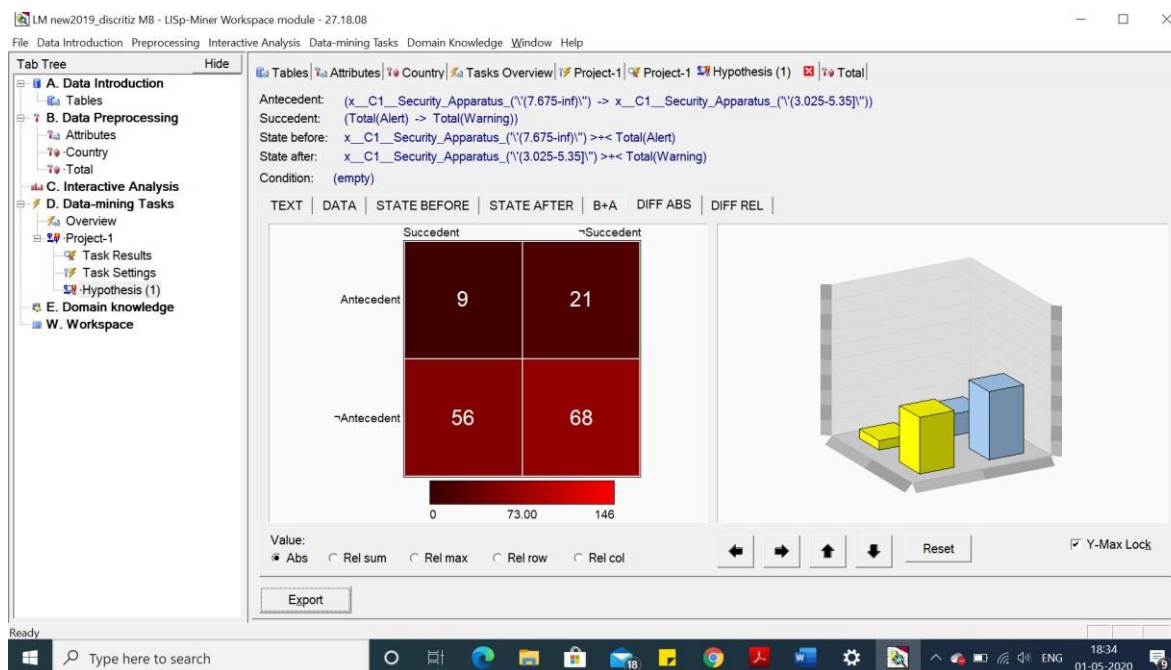
State Before:



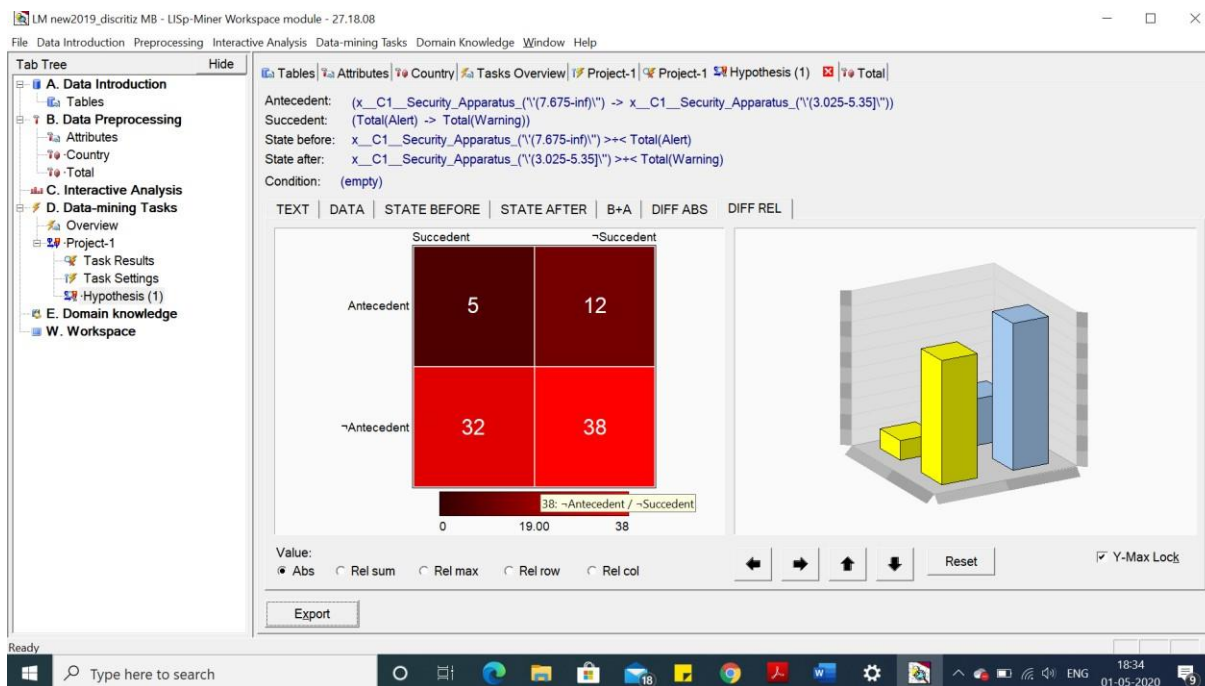
State After:



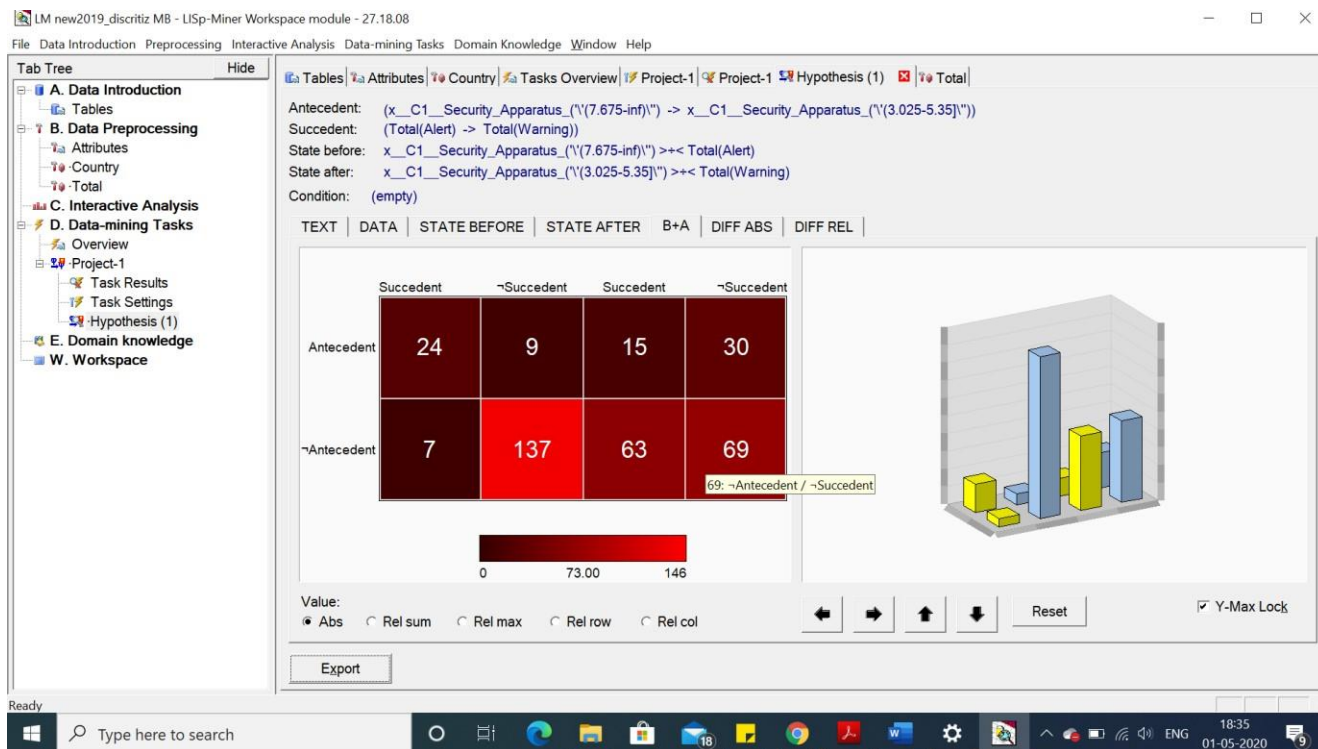
Diff-ABS:



Diff-REL:



B+A:



All Hypothesis:

Task: Project-1
Comment: -
Taskgroup: Default group of tasks
Data matrix: new2019_discr
Task type: Ac4ft-Miner

Task run
Start: 30.4.2020 19:09:47 Total time: 0h 0m 0s
Number of verifications: 256
Number of hypotheses: 29 Mode: Standard

Actual group of hypotheses: All hypotheses
Hypotheses in group: 29 Shown hypotheses: 29 Highlighted: 0

Nr.	Id	Df-Conf	B-Conf	A-Conf	Hypothesis
1	1	0.394	0.727	0.333	(empty) : (x_C1_Security_Apparatus_('\(7.675-inf)\')) -> x_C1_Security_Apparatus_('\(3.025-5.35]\')) >+< (empty) : ^
2	20	0.286	1.000	0.714	(empty) : (x_S1_Demographic_Pressures_('\(7.75-inf)\')) & x_S2_Refugees_and_IDPs_('\(7.775-inf)\')) -> x_S1_C
3	25	0.260	0.944	0.684	(empty) : (x_S2_Refugees_and_IDPs_('\(7.775-inf)\')) & x_X1_External_Intervention_('\(7.675-inf)\')) -> x_S2_Ref
4	27	0.241	0.944	0.704	(empty) : (x_S2_Refugees_and_IDPs_('\(7.775-inf)\')) & x_X1_External_Intervention_('\(7.675-inf)\')) -> x_S2_Ref
5	11	0.213	0.655	0.443	(empty) : (x_E1_Economy_('\(7.65-inf)\')) -> x_E1_Economy_('\(3.35-5.5]\')) >+< (empty) : (Total(Alert) -> Total(War
6	28	0.191	0.647	0.457	(empty) : (x_X1_External_Intervention_('\(7.675-inf)\')) -> x_X1_External_Intervention_('\(3.025-5.35]\')) >+< (empty
7	13	0.174	0.739	0.565	(empty) : (x_E1_Economy_('\(7.65-inf)\')) & x_S1_Demographic_Pressures_('\(7.75-inf)\')) -> x_E1_Economy_('\(
8	15	0.160	0.739	0.579	(empty) : (x_E1_Economy_('\(7.65-inf)\')) & x_S1_Demographic_Pressures_('\(7.75-inf)\')) -> x_E1_Economy_('\(
9	24	0.119	0.769	0.650	(empty) : (x_S2_Refugees_and_IDPs_('\(7.775-inf)\')) -> x_S2_Refugees_and_IDPs_('\(5.55-7.775]\')) >+< (empty)
10	3	0.115	0.941	0.826	(empty) : (x_C1_Security_Apparatus_('\(7.675-inf)\')) & x_E1_Economy_('\(7.65-inf)\')) -> x_C1_Security_Appar
11	18	0.109	0.609	0.500	(empty) : (x_S1_Demographic_Pressures_('\(7.75-inf)\')) -> x_S1_Demographic_Pressures_('\(3.25-5.5]\')) >+< (en
12	4	0.108	0.941	0.833	(empty) : (x_C1_Security_Apparatus_('\(7.675-inf)\')) & x_E1_Economy_('\(7.65-inf)\')) -> x_C1_Security_Appar
13	21	0.083	1.000	0.917	(empty) : (x_S1_Demographic_Pressures_('\(7.75-inf)\')) & x_S2_Refugees_and_IDPs_('\(7.775-inf)\')) -> x_S1_C

INFERENCES FROM RULES

$(x_C1_Security_Apparatus_('\(7.675-inf)\')) \rightarrow x_C1_Security_Apparatus_('\(3.025-5.35]\')) >+< (Total(Alert) \rightarrow Total(Warning))$

$(x_C1_Security_Apparatus_('\(7.675-inf)\')) \rightarrow x_C1_Security_Apparatus_('\(5.35-7.675]\')) >+< (Total(Alert) \rightarrow Total(Warning))$

$(x_E1_Economy_('\(7.65-inf)\')) \rightarrow x_E1_Economy_('\(3.35-5.5]\')) >+< (Total(Alert) \rightarrow Total(Warning))$

- A shift from a factor of 8 plus to 3 or 8 plus to 5 from security Apparatus would transform a country from an alert to a Warning state. It means higher security is slightly increased, resulting in a less fragile state.
- As the economy decreases from 8 plus to 3 the fragility result of country also decreases and hence transforming the country from an alert to a warning state

$(x_S2_Refugees_and_IDPs_('\(7.775-inf)\')) \& x_X1_External_Intervention_('\(7.675-inf)\')) \rightarrow x_S2_Refugees_and_IDPs_('\(3.325-5.55]\')) \& x_X1_External_Intervention_('\(3.025-5.35]\')) >+< (Total(Alert) \rightarrow Total(Warning))$

$(x_S2_Refugees_and_IDPs_(\backslash(7.775-inf)\backslash) \ \& \ x_X1_External_Intervention_(\backslash(7.675-inf)\backslash) \rightarrow x_S2_Refugees_and_IDPs_(\backslash(5.55-7.775]\backslash) \ \& \ x_X1_External_Intervention_(\backslash(5.35-7.675]\backslash)) >\div< (Total(Alert) \rightarrow Total(Warning))$

$(x_X1_External_Intervention_(\backslash(7.675-inf)\backslash) \rightarrow x_X1_External_Intervention_(\backslash(3.025-5.35]\backslash)) >\div< (Total(Alert) \rightarrow Total(Warning))$

$(x_X1_External_Intervention_(\backslash(7.675-inf)\backslash) \rightarrow x_X1_External_Intervention_(\backslash(5.35-7.675]\backslash)) >\div< (Total(Alert) \rightarrow Total(Warning)) \quad 0.63$

- Refugees and internally displaced persons, a higher influx of refugees and IDPs will result in fragility and a decrease in this rating will result in less fragility.
- A significant shift from a factor of 8 plus to 3 from external interventions would transform a country from an alert to a Warning state.

All the other rules are included in other document named – Rules Extracted

CONCLUSION

The data with additional features was collected through the web from different sources and information was pre-processed, discrete and grouped using WEKA for classification.

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. We have discovered and studied various action rules from the dataset out and have action rules explained in this report.

REFERENCES:

- [1] <https://tradingeconomics.com/country-list/inflation-rate>
- [2] [https://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(nominal\)](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal))
- [3] <https://www.kaggle.com/kerneler/starter-countries-database-sqlite-bdd8e6f2-0/data>
- [4] <https://ourworldindata.org/>
- [5] <https://data.worldbank.org>
- [6] <http://hdr.undp.org/en/data>
- [7] <http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/BayesNet.html>
- [8] <http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/JRip.html> [9] <http://fsi.fundforpeace.org/>