**KNOWLEDGE DISCOVERY IN**

**DATABASES**

PROJECT REPORT

ON

Analysis of Fragile States Data

By

BHARGAV NALLANI CHAKRAVARTHULA (801031339)

**CONTENTS**

|  |  |
| --- | --- |
| **S.NO.** | **Topic** |
| 1. | Introduction |
| 2. | Fragile State Index |
| 3. | FSI Indicators |
| 3.1 | FSI Dataset extension |
| 3.2 | New Features of FSI Dataset |
| 4. | Data Extraction |
| 5. | Data Discretization using WEKA |
| 6. | Classification using WEKA |
| 7. | Action Rule Discovery using LISP Miner |
| 8. | Conclusion |
| 9. | References |

# **Introduction**

The aim of the project is to accomplish action rules for the given Fragile State Index Dataset [1] by adding 6 new features and check how these action rules change year by year. The action rules that show change from alert to warning are observed. By using Total as decision feature, the data is discretized and classified as alert, stable and sustainable. Basically, Action rule mining works on a decision system that consists of objects having classes of attributes such as Stable, flexible and decisions.

# **Fragile State Index (FSI)**

The Fragile States Record (FSI) is one of the most vital tools for highlighting both ordinary and critical pressures which a state can experience. The Fund for Peace has gathered country-wise patterns to create this file which assesses existing and up and coming political, financial & social dangers that may challenge the globalized populace. FSI is a yearly positioning of 179 nations over globe based on the Fund for Peace’s restrictive Conflict Assessment System Tool (CAST) analytical platform. It is fundamentally vital to perform prescient investigation and build up patterns in this dataset to recognize and pro-actively react to powerless, falling flat, unsteady, delicate states.

The FSI is the entirety of scores of twelve highlights (to be specific, Demographic pressures, Refugees and IDPs, Group Grievance, Human Flight and Brain Drain, Economic Inequality, State Legitimacy, Public Services, Human Rights, Security Apparatus, Economy, Factionalized Elites, External Intervention) demonstrative of state’s stability. These highlights are broadly categorized into social, financial & political bunches. Each highlight has its esteem characterized inside the span of 1 and 10, with a higher esteem illustrating more delicacy, hence making a scale extending 0−120. Each of these highlights is clarified next sub-section.

# **Indicators**

The 4 indicators in the Fragile state index are

**Social Indicators**

* **Demographic Pressures (S1):** Defined as the burden on the state resulting from both the population and other life preserving properties like food, water etc.
* **Refugees and IDPs (S2):** Defined as the pressure caused upon states caused by shift of population due to social, political, environmental or other reasons.
* **External Intervention (X1):** Defined as influence of external players such as foreign assistance, foreign military intervention, sanctions in the functioning of the state.

**Economic Indicators**

* **Uneven Economic Development (E2):** Defined by inequality in different genres like work, education, financial within the economy, irrespective of actual performance of the economy.
* **Economic Decline (E1):** Defined by the economic decline of the society as a whole as measured by per capita income.
* **Human Flight (E3)**: The economic impact of brain drain (Immigration of educated people who moves out of countries for better opportunity) and the impact this may have on a country’s development

**Cohesion Indicators**

* **Security Apparatus (C1):**  Considers the security threats to a state, such as bombings, terrorist attacks, rebel movements and coups. Measurements include internal conflict, riots/protests, military coups, rebel activity.
* **Factionalized Elites(C2):** Measures power struggles, political competition and transitions, and where elections reliability is on the line.
* **Group Grievance(C3):** Defined by the prevailing division between communities that can impend the country's sovereignty.

**Political Indicators**

* **State Legitimacy(P1):** The Indicator looks at the national’s level of sureness in Government organizations and its functioning.
* **Public Services(P2):** Defined by the availability of delivery of vital services like healthcare, education, sanitation, transportation and others that help the people.
* **Human Rights(P3):** It specifies whether there is extensive exploitation of legal, political and societal rights, counting those of people, groups and establishments

# **Extending the FSI Dataset**

This project proffers finding new designs & patterns utilizing classification of delicacy inside states. We expand this analysis by including six new highlights which may permit hazard evaluation and early caution of struggle. The values for these amplified highlights have been extricated from different websites and included to the unique FSI dataset. The information for four sequential years, i.e., 2012-2015 is compiled and put away as an Excel sheet for further processing & examination.

# **New Features**

The following extra features were added to the original FSI Dataset:

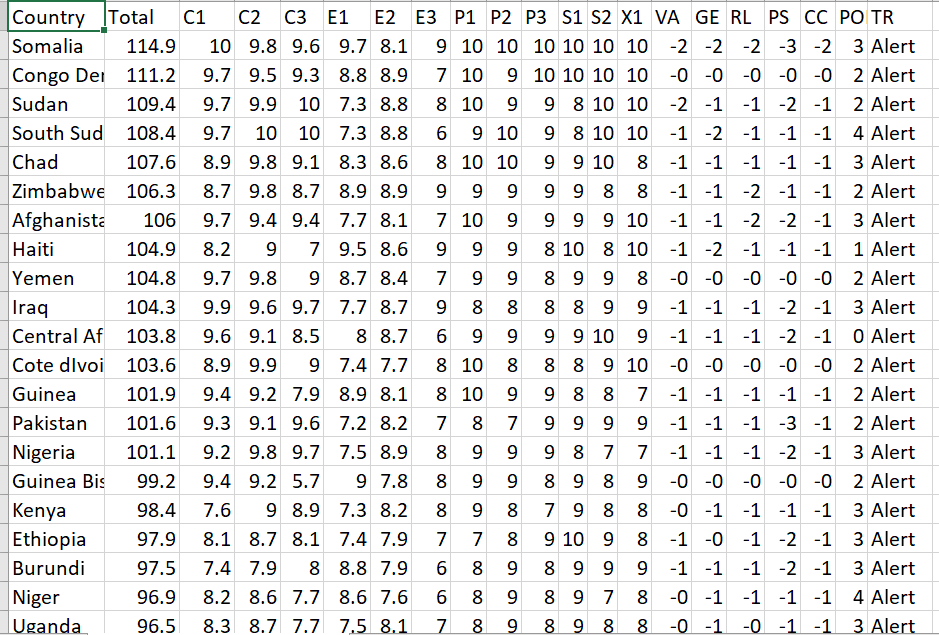
* **Voice and Accountability (VA):** Voice and Accountability refers to the level of acceptable open support in choosing the government & the sum of opportunity to give views & opinion within administration practices are representative of the country’s straightforward, responsible and steady policy.
* **Political Stability (PS):** Political Stability refers to the propensity for a change in the government. It also refers to the violence in the society.
* **Government Effectiveness (GE):** Government effectiveness refers to the policies and practices used by the government to identify the standards of services provided.
* **Rule of Law (RL):** Every state has a Rule of law. Rule of law requires the Government to rule the state with proper establishment of rules, principles and regulations.
* **Population (POP)**: Population refers to the number of people living in a particular state.
* **Control of Corruption (CC):** The Control of Corruption refers to an extent which public power is utilized to make private profits.

1. **Data Extraction**

Different websites from the internet were explored to collect the data for the extended features for all 178 countries. These are listed below:

* http://hdr.undp.org/en/data
* http://info.worldbank.org/governance/wgi/index.aspx#home
* http://economicsandpeace.org

Once the data for 4 consecutive years 2012-2015 was collected and the missing values were calculated, the data was arranged in a Microsoft Excel Sheet.



*Figure 2: Screenshot of extended dataset with existing and new features.*

The complete extended dataset is attached below for reference:



1. **Data Discretization using WEKA**

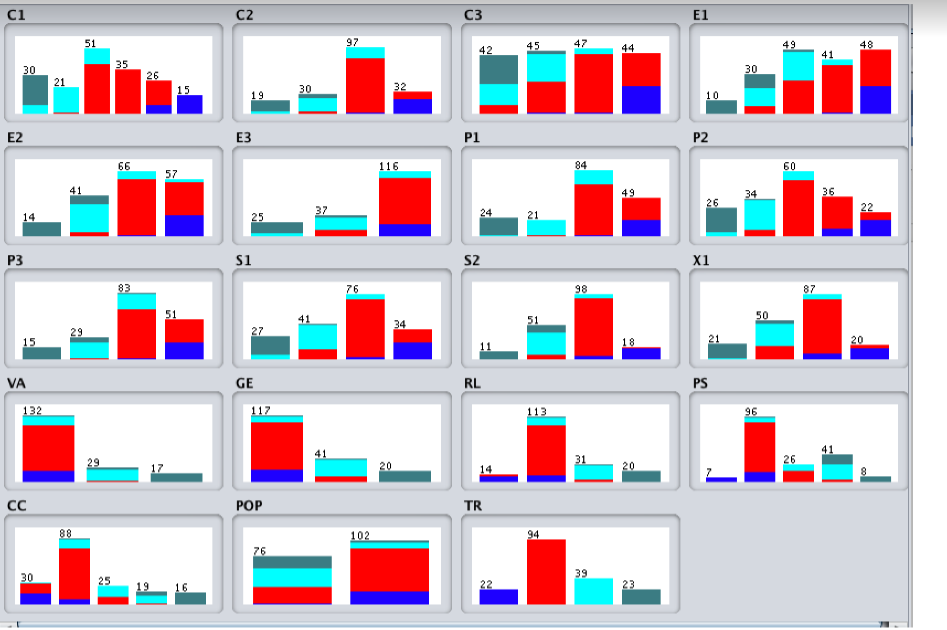
The fundamental point of this project is to find action-rules and to achieve this; the essential errand is to execute classification algorithms which can effectively recognize interesting patterns in the data from which a mining model can be made. Bigger set of numeric values are troublesome to manage with and too numerous learning models do not handle continuous attributes, hence cultivating the need of preprocessing the data. Discretization [3] is a noticeable data reduction strategy which changes over continuous numeric values into limited number of bins (buckets) indicating discrete requested values constituted by a set of intervals, in this way restricting the number of conceivable states. Discretization can be either be done utilizing directed or unsupervised filters. WEKA has two ways to discretize features:

1. the supervised approach utilizing 'multi-interval discretization', (MDL) and
2. the unsupervised approach utilizing straightforward binning. We chose the supervised filters for discretization.

Supervised discretization takes class value into thought and carries out discretization in the taking after three steps:

* Sort the values.
* Place cuts (breakpoints) between values having a place to diverse classes.
* If numerous intervals, consolidate intervals with equal or similar class distributions.

In this project, numeric values of Total are supplanted by the following concepts, Alert, Warning, Stable, Sustainable. Discretization of the dataset is done taking Total as the decision attribute.



*Figure 3: Sample screenshot of the discretized data for all the features derived using WEKA*

The ARFF data files of discretized data for the years 2012-2015 are given below:



1. **Classification Using WEKA**

Classification, pre-processing is performed on data using WEKA. Various decision tree algorithms are used in classification like J48, random forest, random tree etc.

* + **Random Tree**

A Random Tree can be defined as a tree that is drawn randomly from a set of possible trees by taking K random features at each node. We used Random Tree algorithm using a Cross Validation of 10 folds

* + **J48**

J48 is a modified version of C4.5 in WEKA. The C4.5 algorithm generates a classification-decision tree for the given data-set by partitioning the data recursively. The decision is taken using Depth-first strategy.

* + **Random Forest**

Random forest is basically a combination of unpruned classification and regression trees, induced from bootstrap samples of the training data. It uses random feature selection for the process of tree induction. Prediction is made by aggregating the predictions of the unpruned classification and regression trees[3].

The precision results of the classifiers are as given below:

* + **Random Tree:** We used Random Tree algorithm using a Cross Validation of 10 folds. The precision of classifier, for the four years is as follows:

1. 85.9551 % - 2012
2. 93.8202 % - 2013
3. 88.2022 % - 2014
4. 87.0787 % - 2015

**J48 algorithm:** We used J48 algorithm using Cross Validation of 10 folds. The precision of classifier, for the four years is as follows:

* + 1. 84.2697 % -2012
    2. 91.573 % -2013
    3. 88.764 % -2014
    4. 86.5169 % -2015

**Random Forest:** We used Random Forest algorithm using a Cross Validation of 10 folds. The precision of classifier, for the four years is as follows:

* 1. 95.5056 % -2012
  2. 96.6292 % -2013
  3. 92.1348 % -2014
  4. 94.382 % -2015



1. **Action Rule Discovery using LISP Miner**

LISP Miner tool [4] was used for mining of action rules, **Ac4ft-Miner algorithm** was used to generate the action rules. A threshold of 12 was applied which generated the hypothesis, of the overall generated hypothesis the most logical ones were considered. Action rules express which action should be performed to improve the defined state. Action rules suggest a change in behavior that can bring us an advantage. They assume two sets of attributes in the database – stable attributes and flexible attributes. Flexible attributes are the attributes which can be somehow changed by the user (that is, it is possible to change the behavior and in this way to change the attribute). It is to be noted that the following four features have been selected as stable:

* **Demographic Pressure:** As we cannot change the resources of a state at a large amount, we consider this feature as a stable one.
* **Group Grievance:** Group Grievance will be challenging to eradicate and a difficult one to change; hence we are considering this as a stable feature.
* **Country:** Country Name cannot be changed, it is constant at any point of time. We considered country as a stable attribute.
* **External Intervention:** We cannot directly mitigate the alliances with external countries thus intervention will always be there. Hence, we are considering this as a stable feature.
* **Action Rules | Year 2015**

1. (FactionalizedElites(<8.2;10>) & Refugees(<8.2;10>) -> FactionalizedElites(<6.4;8.2)) & Refugees(<4.6;6.4))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

2. (FactionalizedElites(<8.2;10>) & SecurityApparatus(<8.2;10>) & StateLegitimacy(<8;10>) -> FactionalizedElites(<6.4;8.2)) & SecurityApparatus(<6.4;8.2)) & StateLegitimacy(<6;8))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

3. (Economic\_Inequality(<6.4;8.2)) & FactionalizedElites(<8.2;10>) -> Economic\_Inequality(<4.6;6.4)) & FactionalizedElites(<6.4;8.2))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

* **Action Rules | Year 2014**

1. (FactionalizedElites (<8.2;10>) & Refugees(<8.2;10>) -> FactionalizedElites (<6.4;8.2)) & Refugees(<6.4;8.2))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

2. (FactionalizedElites(<8.2;10>) & SecurityApparatus(<8.2;10>) & State\_Legitamacy(<8.2;10>) -> FactionalizedElites (<6.4;8.2)) & SecurityApparatus (<4.6;6.4)) & State\_Legitamacy(<6.4;8.2))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

3. (Economic\_Inequality(<6.4;8.2)) & FactionalizedElites (<8.2;10>) -> Economic\_Inequality(<4.6;6.4)) & FactionalizedElites (<6.4;8.2))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

* **Action Rules | Year 2013**

1. ((Economic\_Inequality(<8.2;10>) & FactionalizedElites (<8.2;10>) -> Economic\_Inequality(<6.4;8.2)) & FactionalizedElites (<4.6;6.4))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

2. (FactionalizedElites (<8.2;10>) & SecurityApparatus (<8.2;10>) & State\_Legitamacy (<8.2;10>) -> FactionalizedElites (<6.4;8.2)) & State\_Legitamacy (<6.4;8.2)) & SecurityApparatus (<6.4;8.2))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

3. (FactionalizedElites (<8.2;10>) & Refugees(<8.2;10>) -> FactionalizedElites (<6.4;8.2)) & Refugees(<4.6;6.4))) **DECISION**-🡪 (Status(Alert) -> Status(Warning))

* **Action Rules | Year 2012**

1. (FactionalizedElites(<8.2;10>) & Rufugees(<8.18;10>) -> FactionalizedElites(<6.4;8.2)) & Rufugees(<2.72;4.54))) **DECISION**-🡪 (Total(Alert) -> Total(Warning))

2. (FactionalizedElites(<8.2;10>) & SecurityApparatus(<8.2;10>) & State\_Legitamacy (<8.08;9.9>) -> FactionalizedElites(<6.4;8.2)) & SecurityApparatus(<4.6;6.4)) & State\_Legitamacy (<6.26;8.08))) **DECISION**-🡪 (Total(Alert) -> Total(Warning))

3. (Economic\_Inequality(<7.54;9.1>) & FactionalizedElites(<8.2;10>) -> Economic\_Inequality(<5.98;7.54)) & FactionalizedElites(<6.4;8.2))) **DECISION**-🡪 (Total(Alert) -> Total(Warning))



**8. Conclusion**

The action rules that were extracted have been presented in the previous sections. There is a common factor in the action rules generated for all four years in the sense that they point towards growing or declining of the values of certain features. Decrease in the range of Factionalized Elites and Security Apparatus have changed the fragility of the state towards the good. Thus, the action rules generated from the decision system can be provided as a predominant tool for countries to adopt a great strategy to maintain stability of the country.

**9. References**

[1] <http://fsi.fundforpeace.org>

[2] Fabrice Muhlenbach and Ricco Rakotomalala (2005). “Discretization of Continuous Attributes”. Encyclopedia of Data Warehousing and Mining, Wang, J. (Ed.), Idea Group Inc., 397-402.

[3] Zhao, Y., & Zhang, Y. (2008). “Comparison of decision tree methods for finding active objects”, Advances in Space Research. 41, 1955–1959.

[4] https://lispminer.vse.cz