

# Classification based on Empathy level by Mining Economic Prosperity and Environmental Indicators

<sup>1</sup>Muhammad Shahbaz, <sup>2</sup>Masooma Iftikhar, <sup>1</sup>Rafia Mahmood

<sup>1</sup>Department of Computer Science & Engineering University of Engineering & Technology

<sup>2</sup>Institute of Business Administration University of the Punjab  
Lahore, Pakistan

## Abstract

*The role of economic prosperity and environmental empathy indicators affects empathy level of a country and shows how much a country is empathetic. We put forward a methodology to classify a country's empathy by dividing it into five labels: (1) apathetic; (2) less-empathetic; (3) empathetic; (4) more-empathetic; (5) future-empathetic, by mining the selected economic prosperity and environmental empathy indicators. We justified our indicators by means of unsupervised clustering, and clustered world countries with respect to the economic prosperity and environmental empathy indicators and assigned those labels. Furthermore, building hierarchical structured decision trees on the labeled dataset helped us in finding important patterns in classification of empathy and evaluating the level of empathy of a country in an interesting and effective way.*

## 1. Introduction

The concept of empathy is explained in New Dictionary of Cultural Literacy by Hirsch et. al., 2002, as, “Empathy is the process of identifying oneself completely with an object or person, sometimes even to the point of responding physically, as when, watching a baseball player swing at a pitch, one feels one's own muscles flex”[1]. An American psychologist Edward Titchener in 1909 interpreted empathy with a German word *Einfühlung*, whereas Titchener, E. B. distinct *Einfühlung* as a “process of humanizing objects, of reading or feeling ourselves into them”[2]. While this German term *Einfühlung* was initially used by Vischer in aesthetics ( In fine art a branch of philosophy deals with expressing beauty and nature) to delegate “the projection of human feeling on to the natural [or physical] world”[3].

As research has been conducted over past hundred years on empathy extending over numerous disciplines including conventional disciplines like philosophy and social, developmental and clinical psychology but also some activist disciplines like Marketing, business administration, business management, sales, civil engineering, construction and engineering education

[4]. Empathy can also be defined as the capacity to recognize and share feelings that are experienced; and act upon those feelings.[1] [2] [3] [4]

## 2. Environmental Empathy

Empathy is categorized in two separate groups:

- i.Environmental empathy
- ii.Human empathy

Environmental empathy is not quantifiable as a country having more awareness, consciousness or concern about their environment conditions is more empathetic. A framework called ten-principle was introduced as General Theory of Sustainability (GTS) that explains how a sustainable society can come up using these principles. This theory was based on a psychological concept but was an important one, which is now popular in other disciplines as well and is the basis of the concept of empathy. It links various issues like eliminating the issue of poverty, proper, careful and supervised usage of the physical environment; by developing, implementing and encouraging empathetic skills in all the countries and implies how sustainable development can be achieved via these issues.

### 2.1. Human Empathy

Methodical and Logical thinking will never be able to grasp all the complexities that lie beneath human nature, the wealth generation, division and allocation processes in fair economies, and other aspects of the world including climatic changes, and patterns, water cycle dynamics, pollution, agricultural growth etc. It will not be possible to resolve the problem “environment versus economy” using methodical thinking alone [4]. Moreover, Empathy requires some individual effort. Most notably, empathy is necessary for the resolving racial oppression, differences and quarrels. Empathy at the personal level can make true equality achievable at the societal level.

We devised a process to categorize a country's empathy level into one of five labels: (1) apathetic; (2) less-empathetic; (3) empathetic; (4) more-empathetic;

(5) future-empathetic to perform a correlation analysis between empathy and economic prosperity, environmental empathy indicators. We performed a unique mechanism of preprocessing for cleaning and smoothing the available data of the indicators for each year and performed year wise cleaning of ten years (2000-2009). After that by using K-mean clustering the World countries were grouped into five clusters initially, the world countries are clustered based on economic prosperity and environmental empathy indicators for every year from 2000 to 2009. The resulted clusters do not have a label so the world countries were grouped into five classes but the fact that a country belongs to a particular group cannot be identified. A method of labeling cluster is defined in second part of section 5 according to which we should find the empathy level of each world country. To classify a country's empathy level into one of five classes. (1) apathetic; (2) less-empathetic; (3) empathetic; (4) more-empathetic; (5) future-empathetic, the correlation analysis has been performed on each year cluster movement of a country from 2000-2009 and to obtain an average clustered value we assign a value to each empathy level of a country for every year and find an average clustered value. We then applied "J48 decision tree classification algorithm" to the cluster to simulate economic prosperity and environmental empathy indicators on a hierarchical structure for ease of its use and tracking when assessing the country's empathy level. Thus, using this technique decision maker should decide which economic prosperity and environmental empathy indicator plays an important role to improve the empathy level of a country.

The paper is organized as follows. Section 2 elaborates the correlation between economic prosperity and environmental empathy indicators with empathy. Section 3 reviews the key literature on clustering and decision tree classification. Section 4 reviews the key literature on Empathy and indicators shown in table 1, 2. We develop a method for clustering world countries with respect to empathy in Section 5. In Section 6, we perform experiments and finally the work concluded in Section 7.

### 3. Economic Prosperity and Empathy

The research work methodology focuses on implementing data mining technology to determine a country's empathy level and the development of knowledge-based models to predict a country's empathy class we have tried to build a relationship of economic and environmental empathy indicators with empathy. Research has found that under developed

countries lack the financial resources necessary to provide the legal safeguards that make possible economic development. Contentment of basic needs, such as survival, may require sacrifice of personal values, such as morality. Park says "scarcity is the father of corruption" [5].

In the same way using economic and environmental indicators (mentioned in table 1 and 2) of a country we can get to know the level of empathy a country might show as these indicators put a direct or indirect effect on empathy. In general, the paper tries to explain that the stable the economy of a country is; the greater is the empathy altitude of the country, which shows the economic dependency on the empathy coefficient.

Globally, as a whole, we need to study how to spark or rouse empathy and concern, and how to maintain and act from empathy and concern. There is a need to study how we can make it politically meaningful. Additionally, we need to find ways to do this that can then advance sustainable expansion and maturity, rather than only focusing on unchecked fossil fuel consumption and polluting industrial processes [6].

#### 3.1. Correlation of Factors

Economic factors, such as GDP, GNI, Gini, Food Production Index, Food and Crop Production Index, Literacy Rate, Public Spending on Education and Tax Rate paid by the people of that country as well as environmental factors such as forest area directly related to empathy. The greater the value of these indicators would be for a country, the country would be economically stable. And economic stability will make people will tend to give more attention to the needy people and show empathy towards them [5]. Some of the indicators such as Inflation Rate, Unemployment, Interest Rate and the environmental factors such as EH, EHEH, EHAIR, EHWATER, EV, EVAIR, EVWATER (given in table 2) are inversely related to the empathy. This is because a greater values of these indicators will show more frustration and stress among the people of that country and economic instability, as a result they will be insensitive to each other [5].

### 4. Clustering and Decision Tree Classification

#### 4.1. Clustering

The clustering problem is defined as the problem of finding groups of data points in a given data set. Each of these disjoint sub collection of data is known as a cluster and can be defined as a region in which the

density of objects is high than in other regions. Clustering is one of the most vital *unsupervised learning* techniques; it deals with finding a *structure* in a collection of unlabeled data. This is the process of organizing objects into groups whose members are similar in some common attribute(s) whereas dissimilar to the objects in other groups. This technique aids in pattern detection using simplification steps. Using this unsupervised learning process, results from every step can use in data concept building.

Clustering has its benefits when the data sets defined and a general pattern needs to be determined from the data. You can create a specific number of groups, depending on your business needs. One important benefit of clustering over classification is that every attribute in the data set will be used to analyze the data. A major disadvantage of using clustering is that the user is required to know ahead of time how many groups he wants to create. For a user without any real knowledge of his data, this might be difficult. Should you create three groups or six groups? It might take several steps of trial and error to determine the ideal number of groups to create.

Nonetheless, clustering can be the most useful data mining technique you can use. It can quickly take your entire set of data and turn it into groups, from which one can quickly make conclusions. If meaningful clusters are the goal, then the resulting clusters should capture the “natural” structure of the data. Usual pattern clustering activity includes the following steps: pattern representation (includes feature extraction and/or selection), designation of a pattern proximity measure appropriate to the data domain, clustering, data abstraction, and assessment of output [7] [8].

Clustering is repeatedly confused with classification, but there are differences between the two. In classification, the objects assigned to predefined classes, whereas in clustering, the classes also be defined and this is done after the clusters are made and according to the requirement of what the analyzer wants. In comparison to clustering, classification can do in two different manners: supervised classification and unsupervised classification. In former one, the user provided with inputs having class labels associated with them whereas in the later one no class labels assigned to the input data. Clustering classified as the later technique of classification as data that grouped together using clustering technique does not bear any labels [9].

**Table 1. Economic Prosperity Indicators**

Indicator Name	Description
I1. GDP	Goods produced per person
I2. Gini Index	Dispersion of wealth
I3. Unemployment	Joblessness
I4. Public Spending on Education	Public spending on education per year
I5. Population Growth	Population growth per year
I6. Literacy Rate	Rate of educated people
I7. Interest Rate	Rate of Interest paid per yr
I8. Inflation	Rate of change of price level
I9. Public Health Expenditure	% health expenditure by Govt.
I10. GNI	Gross National Income
I12. Food Production Index	Country's Food Production
I13. Crop Production Index	Agricultural Production
I14. Tax Rate	Amount of tax paid

**Table 2. Environmental Empathy Indicators**

Indicator Name	Description
I11. Forest Area	Country's Forest Area
I15. CO2 Emission	Quantity of CO2 released to atmosphere
I16. EPI	Environmental Performance Index
I17. EH	Environmental Health
I18. EHAIR	Air Pollution (effects on humans)
I19. EHEH	Environmental burden of disease
I20. EHWATER	Water Pollution (effects on humans)
I21. EV	Ecosystem Vitality
I22. EVAG	Agriculture
I23. EVAIR	Air Pollution (effects on ecosystem)
I24. EVBH	Biodiversity & Habitat
I25. EVCLIMATE	Climate Change

Clustering divides the dataset into defined number of groups. The quality of the clusters determined by the average error, which calculated by the distance of every member from its cluster center using squared error criteria a well-liked clustering method that minimizes the clustering error, is the k-means algorithm. K-means algorithm is explained below briefly in the following steps [10]:

1.  $K$  initial cluster centers are selected randomly from the given  $n$  points  $\{x_1, x_2, \dots, x_n\}$ ,  $k \leq n$ .

Let cluster centers be  $c_1, c_2, \dots, c_k$

2. Assign each point  $x_i$ ,  $i = 1, 2, \dots, n$  to the cluster  $C_j$  corresponding to the cluster center  $c_j$ , for  $j = 1, 2, \dots, k$  iff  $\|x_i - c_j\| \leq \|x_i - c_p\|$ ,  $p = 1, 2, \dots, k$  and  $j \neq p$ .

3. Calculate new cluster centers

$c1^*, c2^*, \dots, ck^*$  as follows  $ci^* = \frac{1}{ni} \sum_{x_j \in ci} x_j$  for  $i = 1, 2, \dots, k$  where  $ni$  is the number of data points belonging to the cluster  $ci$ .

4. If  $ci^* = ci, \forall i = 1, 2, \dots, k$ , then terminate. Otherwise continue from step 2.

It is clear from the step 1 that the clustering results depend on the opening cluster centers, which are chosen arbitrarily [11]. Using the algorithm the datasets are divided into the desired number of clusters and converges to a local minimum [7]. The algorithm does not specify anything about the labels of clusters. Briefly explaining the K means algorithm, it accepts the dataset and the number of clusters made as its input. Initial clusters made from the rows of the data set where  $K$  also denotes the number of clusters made. In the next step, Arithmetic Mean of the cluster is calculated. In each of the first  $K$  initial clusters, there is only one record. K-Means assigns each record of the dataset to only one of the initial clusters made. Every record assigned to the nearest cluster, which is most alike using a measure of distance or similarity like the Euclidean Distance Measure or Manhattan/City-Block Distance Measure. The algorithm then re-assigns each record in the dataset to the most similar cluster and the arithmetic mean of all the clusters in the dataset calculated again. These steps are repeated until stable clusters formed and the procedure is completed. Stable clusters are formed when new iterations of the K-Means clustering algorithm does not create new clusters as the cluster center or Arithmetic Mean of each cluster formed is the same as the previous cluster center. The number of clusters that is generally denoted by  $k$  has to be determined. Clustering techniques used to discover natural groups in data sets, and to identify unknown relation that might reside there, without having any background knowledge of the characteristics of the data. They used in a range of areas including computer vision, data mining, bioinformatics, Web mining, and gene expression analysis [12]. In recent years, it recognized that the partitioned clustering technique is well suited for clustering a large dataset due to their moderately low computational requirements. Behavioral clustering and segmentation help derive tactical marketing initiatives by using the variables that determine customer shareholder value. By showing demographic clustering and segmentation within the behavioral segments, one can define strategic marketing campaigns and choose the appropriate marketing channel and advertising for the tactical campaign. It is then possible to target those customers most likely to exhibit the looked-for behavior by generating predictive models.

## 4.2. Decision Tree Classification

Decision trees are the compact form of the data. Each node in a decision tree represents a feature, and each branch represents a value from its domain values and leaves bear the labels of the data. Instances are classified starting at the root node and sorted based on their feature values. The feature that most excellent divides the training data would be the root node of the tree. Different strategies can be used to find out the best attributes for decreasing the disorder in the data. A dataset contains a certain amount of information and when you select random dataset, it has high entropy. Construction of DTs works towards reducing the amount of entropy in the data and alternatively, increases the amount of information exhibited by the data. In ID3 or any other decision tree's algorithms mostly entropy is used to find out the most appropriate attribute for data division for constructing decision trees [13].

Let  $X$  be the dataset containing  $n$  sample.  $X = \{x_1, x_2, \dots, x_n\}$ . Suppose that the dataset  $X$  is labeled by  $k$  distinct class labels  $(1, 2, \dots, k)$ . Let  $x_i$  be the number of samples of  $X$  having class label  $i$ .

1. Information gain (IG) for any data attribute present in the dataset  $X$  can be calculated as,

$$IG = \sum_{i=1}^n P_i \log_2(P_i)$$

Where  $P_i$  is the probability of the class  $i$  and  $P_i = x_i/x$

$$IG = \sum_{i=1}^n \frac{x_i}{x} \log_2 \left( \frac{x_i}{x} \right)$$

Entropy (E) based on dataset partitions can be calculated by

$$E = \sum_{i=1}^n \frac{x_{1i}}{x} | \dots | \frac{x_{mi}}{x} = IG$$

Net gain (NG) of the attribute will be.

$$NG = (IG - E)$$

The attribute having largest gain value will be used as the node to split the data.

Details of the decision trees algorithm can be found in various papers and/or books on data mining [13]. This algorithm will produce a tree like structure having attributes at the nodes and labels at the leaves [9].

## 5. Related Work

Literature survey provides appraisal of different factors including GDP, Poverty, Tourism, GNI, debt

and Exchange rate, Educational institutions, Environment, donation, health care services etc. that greatly affect the economic prosperity. The influence of these factors can be negative and positive with country's sustainability [14]. Statistical, data mining and machine learning, techniques are used by the researcher to extract association rules and predict future trends of economy prosperity on the bases of above indicators [5].

Empathy is a non-quantifiable indicator that also has sound impact on a country's development. According to A.Guergachi hypotheses a country is said to be prosperous if the individuals of that country are more concerned to protect their environment. According to his proposed framework that based upon ten principles of GTS (General theory of sustainability) showed a close association between protection of environment and economy prosperity. Therefore, people's sensitivity towards climate and economic sustainability are directly proportional. These principles highlight some issues which includes limitation of engineering, industrial methods, poverty, wealth generating economic activities that are responsible for polluted environment and the most important how to develop the sense of empathy in individuals [4]. Nelson, L. S. 2007, focuses on finding ways to promote sustainable development by the act of empathy and compassion. People having positive attitudes are more eager to help others in tough time. Positive acts in the environment can influence corporations, nongovernmental organizations, communities, and individuals by raising the sense of empathy and producing positive emotions so that they can actively participate in funding to reduce the poverty and cure disease [6]. The research work done by Zanakis, S. H. and I. Becerra-Fernandez, 2005, showed the use of data mining and different statistical techniques to identify the essential factors associated with a country's competitiveness and the development of knowledge discovery in databases model to predict it. The dataset that used to predict competitiveness score consists of 55 independent variables and 43 countries listed in the '99 WCY (World Competitive Yearbook). The 55 variables like economic, internationalization, governmental, financial, infrastructure, management, science and technology, as well as demographic and cultural characteristics were included in dataset. For analyzing data techniques i.e. stepwise regression, weighted non-linear programming, artificial neural networks, and classification and regression trees are used. The analysis of established priorities for improving country competitiveness showed that the country with higher gross domestic investment, savings and private consumption, more imports of goods and services than export and large productivity is prosperous [15]. We

have considered all the possible influences on well-being that have identified in the literature. Indicators that possibly influence the well-being falls under seven headings (1) income; (2) personal characteristics; (3) social characteristics; (4) time utilization; (5) attitudes towards self, life and others; (6) relationships; and (7) the economic, social and political environment of the country they are living in. Comparative studies on inequality have suffered from methodological problems, which mean that issues related to the causes of cross-national variation in inequality remain unresolved. By examining new data on social policy institutions and income inequality among different population groups, study provides basis for evaluating different, and divergent, theories on the welfare state and equality. There are three cases of illustrations: family policy and child poverty, unemployment benefits and poverty among working aged, and old age pensions and poverty among the elderly [16].

Crouch, G. I. and J. Ritchie, examined associations between society's prosperity and tourism using a competitiveness framework on the bases of different factors (culture, physical, social values and norm, concern towards environment etc.) [17]. According to author, improved quality life shows that society is prosperous. Factors including GDP, GINI index, Poverty, literacy rate, tourism is also participates to measure the society prosperity. The use of knowledge discovery techniques in analysis showed, the nation that are more concerned towards their environment and maintenance of their heritage (like historical places), increased their GNI and GDP 10 to 15% per year by attracting more tourist. Results showed tourism development entails economic prosperity [17]. The research carried out by Guillaumont Jeanneney, S. and P. Hua, 2001, showed the influence of the real exchange rate on urban/rural per capita real income inequality by analyzing yearly data (1982 to 1994) from 28 provinces of China. The reason found was the vast spasm between the rates of production of goods in metropolitan than countryside areas. This leads to the large difference between the urban and rural economy growth. However, this variation alleviated by certain factors including price ratio, real public wages ratio of industrial/agricultural and protection rates. The exchange rate difference totally vanished from coastal areas due to the rural industrial activities [18]. Existing literature concluded although, a lot of research has been done to measure economic prosperity on the bases of different factors but still there is lack of work that classify economic sustainability with respect to empathy along with these indicators. Our findings by using K- means clustering (Data mining approach) show the classification of economic prosperity with respect to empathy under the influence of 25

indicators. Next section describes proposed methodology.

## 6. Proposed Methodology

Twenty five (25) economic prosperity indicators and the environmental empathy coefficient are selected (as show in table 1 and table 2) which are quantifiable and data is available at “The world Bank” and “Yale Center for Environmental Law & Policy”[19] [20].

Initially we collected data for 213 world countries over a span of ten years (2000-2009) for each indicator. The data is carefully analyzed and it was concluded that data of only 58 countries is complete enough to be used. Let  $WC_n$  represents countries where  $n= 1, 2, \dots, 58$ . We applied K-Mean Clustering to group the similar countries together to achieve high intra-cluster similarity and low inter-cluster similarity between WC [9].

### 6.1.Preprocessing

In order to fill in missing values, smooth out noise and correct inconsistencies in data we proposed five step preprocessing mechanism for cleaning and smoothing the data for our work.

**Step1:** Ignore the tuple of WC in which no data available against  $I_n$

**Step2:** Ignore the economic prosperity and environmental empathy indicator  $I_n$  for which no data available against  $WC_n$

**Step3:** Ignore the tuple (by setting a threshold of 30%) of WC in which less than 30% data is available against the  $I_n$

**Step4:** Fill in the missing values by using trend analysis and moving averages of WC against  $I_n$

**Step5:** Normalize the data by using the below equation, Eq. (1)

$$X / (I_{max} - I_{min}) \tag{1}$$

Where, X = observed value of the indicator of that year  
 $I_{max}$  = Maximum value of the indicator of that year  
 $I_{min}$  = Minimum value of the indicator of that year

### 6.2. Labeling Cluster

Let  $C_1, C_2, C_3, C_4, C_5$  represents cluster centers.  $WC_n$  are world countries we want to classify by using k-mean clustering algorithm. As dataset has no labels which lead us no option but to do unsupervised classification.

To select unique belonging of a world country, WC to any of the clusters, we used the “frequent membership rule” as illustrated in Eq. (2)

$$\forall WC \in WC_n \text{ and } C_i \in C_x \therefore WC \in C_i$$

$$iff [\text{membership}(WC \in C_i)] > [\text{membership}(WC \in C_n)] \tag{2}$$

Where  $n = 1, 2, 3, 4, 5$  and  $i =$  the number of cluster [9].

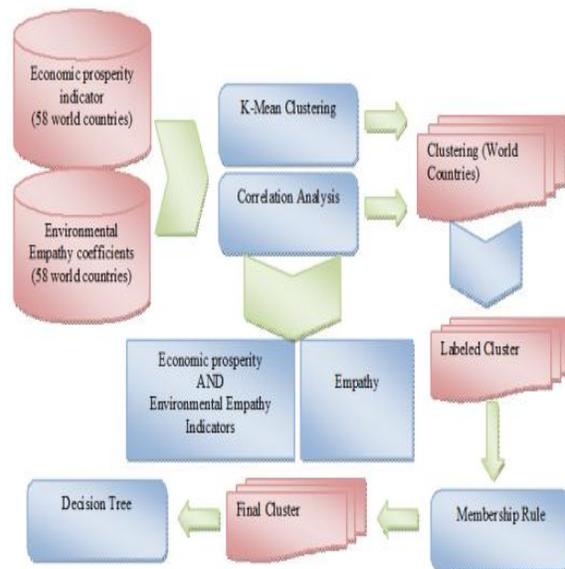


Figure 1. Proposed Methodology

### 6.3. Decision tree

For better visualization of the data and to get to know which indicator provides the greatest information gain, we then built DTs from the preprocessed data and their attributes. This brought us on the decision that which indicator puts the most effect on empathy, as well as removed entropy in the results as this data was randomly collected. The decision tree will enable researchers to relate individual empathy indicators to improve economic development criterions for a given country.

## 7. Result and Discussion

The data of 25 indicators of economic prosperity and environmental empathy of 58 world countries is stored in the database, a skeleton of which is shown in the table 3.

**Table 3. Orientation of Economic Prosperity and Environmental Empathy Indicators Database**

World Countries	2000... 2009				
	$I_1$	$I_2$	$I_3$	...	$I_n$
$WC_1$	$I_1 WC_1$	$I_2 WC_1$	$I_3 WC_1$	.....	$I_n WC_1$
$WC_2$	$I_1 WC_2$	$I_2 WC_2$	$I_3 WC_2$	.....	$I_n WC_2$
$WC_3$	$I_1 WC_3$	$I_2 WC_3$	$I_3 WC_3$	.....	$I_n WC_3$
.....	$I_i WC_i$	.....	.....	.....	.....
$WC_n$	$I_1 WC_n$	$I_2 WC_n$	$I_3 WC_n$		$I_n WC_n$

By using K-mean clustering the countries of the world are grouped into five clusters, the world countries (WCs) are clustered on the basis of prosperity and environmental empathy indicators for every year starting from 2000 till 2009. The country can be represented as follow:

$$X (I_1, I_2, I_3, I_4, I_5, I_6... I_{25})$$

Where X is world country name and I represent the economic prosperity and environmental empathy indicators. The proposed methodology is representing in Figure 1. A method of labeling cluster is defined in part second of section 5 according to which we should find the empathy level of each world country. To classify a country's empathy level into one of five classes, the correlation analysis performed and to obtain an average clustered value we assign a value to each empathy level of a country for every year as follows:

- Apathetic = 0
- Less empathetic = 1
- Empathetic = 2
- More empathetic = 3
- Future empathetic = 4

By using the formula of calculating average shown in Eq. (3) the average cluster value of a country has been obtained

$$Avg = \sum_{i=1}^n \frac{C_i}{n} \quad (3)$$

Average value of a cluster of a world country obtained which shown in the Table 4. As a result, the unsupervised data clustered and now we can assign the Class Label to each country's empathy level:

- Cluster 1 ↔ Apathetic
- Cluster 2 ↔ Less empathetic
- Cluster 3 ↔ Empathetic
- Cluster 4 ↔ More empathetic
- Cluster 5 ↔ Future empathetic

In Figure 2 the clusters plotted in such a way that classes of empathy drawn on the x-axis and the period of empathetic era is plotted at y-axis and tendency of temporal movement of world countries from one class to another shown. The critical level of individual indicator does not reflect by cluster so a decision tree is built to evaluate the empathy level of a country as shown in Figure 3 [9].

**Table 4. Classification of World Countries w.r.t Economic Prosperity and Environmental Empathy Indicators**

Apathetic	Less Empathetic	Empathetic	More Empathetic	Future Empathetic
Angola Congo Dominica India Nigeria Pakistan Sudan Yemen	Paraguay	Algeria Argentina Armenia Bolivia Brazil Colombia Costa Rica Ecuador Egypt El Salvador Georgia Indonesia Iran Kyrgyz Latvia Mexico Moldova Mongolia Panama Per Philippines Romania Syria Thailand Ukraine Uruguay Vietnam	Kazakhstan Kuwait Libya Oman Russia Venezuela	Australia Bulgaria Canada Croatia Denmark Estonia Hungary Lithuania Malaysia Norway Poland Qatar Saudi Arabia UAE UK USA

The attribute with the highest information gain (i.e.,  $I_{19}$ ) will be used as the root node. If  $I_{19} \leq 34.051365$ , then the empathy level of a country will be Apathetic. If the value of  $I_{19} > 34.051365$ , the attribute with the highest information gain ( $I_{15}$ ), will be evaluated, which results into Apathetic category or class. If the value of  $I_{15} > 4.368517$ , the attribute with third highest value of information gain (i.e.,  $I_{10}$ ) will be evaluated and then the empathy level of a country is classified as Empathetic.

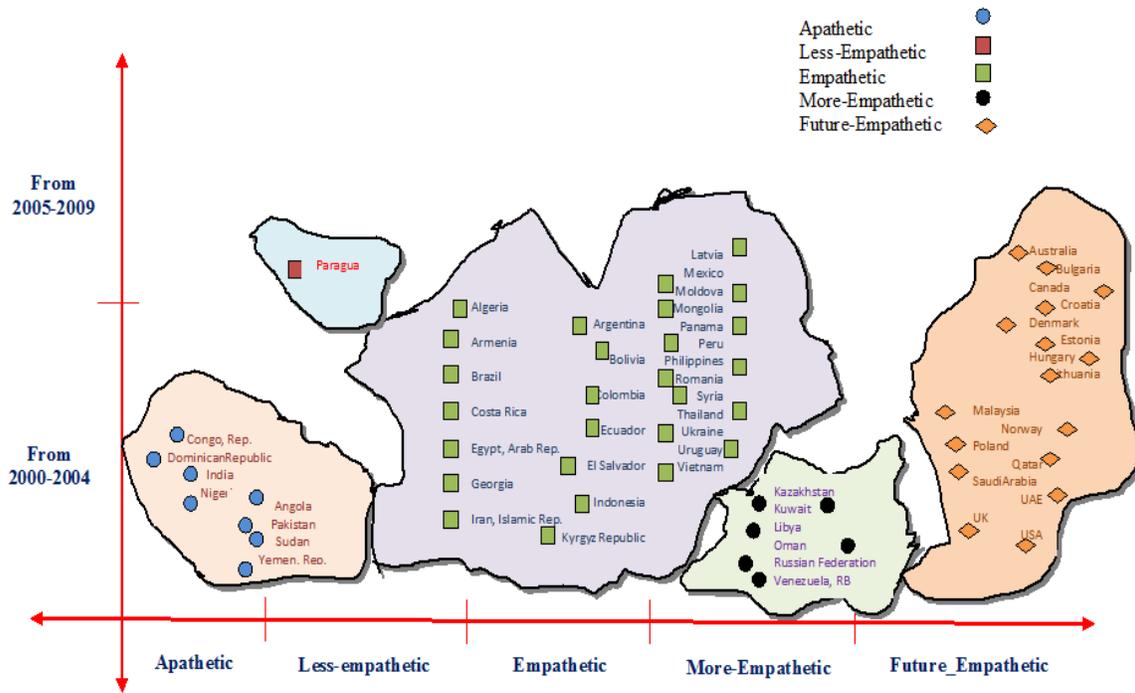


Figure 2. Labeled Clusters of Countries of the World w.r.t Empathy Level

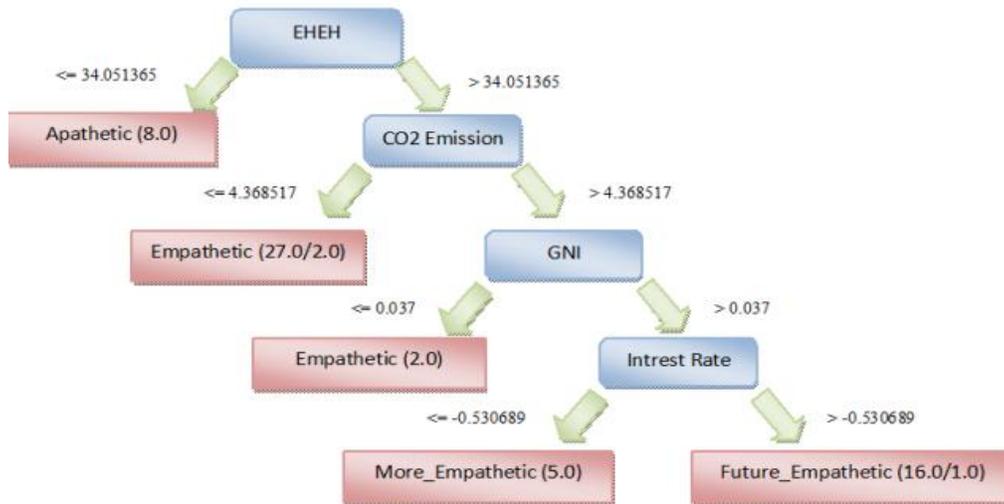


Figure 3. J48 Decision Tree (weka 3.6.8)

Similar mechanism is adopted to traverse the whole tree. In order to deploy these results we only need to control the values of I19, I15, I10 and I7 indicator to make country empathetic in future and classify the level of empathy exist in a country.

The analysis is done using WEKA (open source available free of cost). The detailed summary of the analysis shown below

```

=== Summary ===
Correctly Classified Instances      44      75.8621 %
Incorrectly Classified Instances    14      24.1379 %
Kappa statistic                    0.6383
Mean absolute error                0.1082
Root mean squared error            0.2983
Relative absolute error            39.1767 %
Root relative squared error        80.6459 %
Total Number of Instances          58
    
```

```

=== Confusion Matrix ===
  a    b    c    d    e ← classified as
23    0    3    1    0  a = empathetic
 1    7    0    0    0  b = apathetic
 2    0   12    2    0  c = future_empathetic
 0    0    4    2    0  d = more_empathetic
 1    0    0    0    0  e = less_empathetic
    
```

**Table 5. Detailed Accuracy by Class**

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.852	0.129	0.852	0.852	0.852	0.863	Empathetic
	0.875	0	1	0.875	0.933	0.938	Apathetic
	0.75	0.167	0.632	0.75	0.686	0.756	future_empathetic
	0.333	0.058	0.4	0.333	0.364	0.545	more_empathetic
	0	0	0	0	0	0.307	less_empathetic
<b>Weighted Avg.</b>	<b>0.759</b>	<b>0.112</b>	<b>0.75</b>	<b>0.759</b>	<b>0.752</b>	<b>0.801</b>	

## 8. Conclusion

The data of 25 indicators of economic prosperity and environmental empathy of 58 world countries are mined to classify the empathy level of country into five classes: (1) apathetic; (2) less-empathetic; (3) empathetic; (4) more-empathetic ;( 5) future-empathetic to evaluate the effect of these indicators on empathy and predict which country have what level of empathy. The results also guide us if these indicators value will be changed then what will happen to it empathetic attitude. These results also give controlling parameters that how a country can be shifted from Apathetic to future empathetic.

## 9. References

[1] E. Hirsch Jr, *et al.*, *The new dictionary of cultural literacy*: Download iTunes eBook, 2002.

[2] E. B. Titchener, *A text-book of psychology*: Macmillan, 1921.

[3] G. W. Pigman, "Freud and the history of empathy," *The International Journal of Psychoanalysis*, 1995.

[4] D. A. Swayne, *et al.*, "Empathy: a Unifying Approach to Address the Dilemma of Environment versus Economy'," 2010.

[5] G. R. Franke and S. S. Nadler, "Culture, economic development, and national ethical attitudes," *Journal of Business Research*, vol. 61, pp. 254-264, 2008.

[6] L. S. Nelson, "Inquiry, empathy and art in the social sciences: Thinking and feeling like a social artist," *The Social Science Journal*, vol. 44, pp. 3-9, 2007.

[7] J. Gao and D. B. Hitchcock, "James–Stein shrinkage to improve k-means cluster analysis," *Computational Statistics & Data Analysis*, vol. 54, pp. 2113-2127, 2010.

[8] S. Pande, *et al.*, "Data Clustering Using Data Mining Techniques," 2012.

[9] M. Shaheen, *et al.*, "Mining sustainability indicators to classify hydrocarbon development," *Knowledge-Based Systems*, vol. 24, pp. 1159-1168, 2011.

[10] S. Bandyopadhyay and U. Maulik, "An evolutionary technique based on K-means algorithm for optimal clustering in RN," *Information Sciences*, vol. 146, pp. 221-237, 2002.

[11] D. Reddy and P. K. Jana, "Initialization for k-means Clustering using Voronoi Diagram," *Procedia Technology*, vol. 4, pp. 395-400, 2012.

[12] M. Teboulle, *et al.*, "Clustering with entropy-like k-means algorithms," in *Grouping Multidimensional Data*, ed: Springer, 2006, pp. 127-160.

[13] J. Han and M. Kamber, "Data mining: Concepts and techniques," *China Machine Press*, vol. 8, pp. 3-6, 2001.

[14] M. R. Andreychik and M. J. Gill, "Do negative implicit associations indicate negative attitudes? Social explanations moderate whether ostensible "negative" associations are prejudice-based or empathy-based," *Journal of Experimental Social Psychology*, vol. 48, pp. 1082-1093, 2012.

[15] S. H. Zanakis and I. Becerra-Fernandez, "Competitiveness of nations: A knowledge discovery examination," *European Journal of Operational Research*, vol. 166, pp. 185-211, 2005.

[16] J. Palme, "Welfare states and inequality: Institutional designs and distributive outcome," *Research in Social Stratification and Mobility*, vol. 24, pp. 387-403, 2006.

[17] G. I. Crouch and J. Ritchie, "Tourism, competitiveness, and societal prosperity," *Journal of Business Research*, vol. 44, pp. 137-152, 1999.

[18] S. Guillaumont Jeanneney and P. Hua, "How does real exchange rate influence income inequality between urban and rural areas in China?," *Journal of Development Economics*, vol. 64, pp. 529-545, 2001.

[19] TheWorldBankGroup. (2013, The World Bank Group Data Indicators <<http://data.worldbank.org/indicator>>. Available: <http://data.worldbank.org/indicator>

[20] YaleUniversity, "Environmental Performance Index <<http://epi.yale.edu/downloads>>," 2013.