# Format Conversions for Open Source Data Mining Implementation in Digital Economy Ranking

Mădălina ZURINI
Academy of Economic Studies, Bucharest, Romania
madalina.zurini@gmail.com

*The terms of digital economy and e-readiness are introduced and presented in the present area of evolution of economy stage within Europe and not only. The main aspects treated in the European University Institute ranking for e-readiness are briefly covered. An open source convertor for excel to arff file format is presented and used in the purpose of input transformation for using Weka Data Mining, an open source tool. The framework of ExceltoArff tool is demonstrated in a practical use, the one of 70 countries registrations of EUI Ranking.*

***Keywords:*** *Conversion Process, Open Source, Digital Economy, Clustering Analysis, EUI 2010 Ranking*

## 1 Digital economy ranking

Knowledge society represents a new stage in human evolution, a superior quality lifestyle that involves intensive use of IT in all spheres of human activity, with major social and economic changes. Democracy, communication, understanding and cooperation are the main characteristics of this society, which makes knowledge society to be based on the multitude resources offered by Internet access.

Digital economy is defined in [2] by the changing characteristics of information, computing and communication, transforming it into the driver of economic growth and social change. Since 1998, the US Department of Commerce issued The Emerging Digital Economy, where it was recognized the accelerating importance of the Internet and e-commerce. There is considered to be a strong relation between e-commerce and the digital economy itself. Major studies revealed the fact that digital economy is widely spread within e-commerce.

E-readiness, as seen in [1], is the ability of a country to use information and communication technologies in order to develop its economy. Starting from the need of measuring the level of e-readiness along the countries from all over the globe, UNPAN, World Bank and Economist Intelligence Unit first calculated macro indicators, called e-readiness indices, which helped in evaluating the impact given by ITC within a specific territory.

Since 2000, the world's largest economies were introduced in a global study regarding the impact given by the use of ITC upon the consumers, businesses and governments.

In the EUI e-Readiness ranking, first called, later Digital Economy Ranking, over 100 criteria, qualitative and quantitative, are evaluated for each country from the list of 70. Each criteria is scored, aggregated within every category defined, and also as o total, an average given as a whole. The categories available are:

- *connectivity and technology infrastructure* is associated to the need of reliable, convenient and affordable access to voice and data services;
- business environment is the ranking indicator which evaluates the general business climate;
- social and cultural environment is the characteristic analyzed along with the connectivity, the access;
- legal environment refer to the overall legal framework and specific laws regarding Internet use;
- government policy and vision is the indicator for the evaluation of the overall government area and the ability of it to lead towards a digital future;

- consumer and business adoption is the indicator which refer to the actual utilization of digital channels by people and the business companies.

All data presented in the following report are extracted from the Digital economy ranking 2010, available at [5]. In table 1, the six characteristics with which 70 countries are being classified in EUI e-Readiness ranking, along with their subcategories are presented in Table 1. For each subcategory and also category a weight is given used for the ranking aggregation.

**Table 1.** Categories and subcategories weights

| **No.** | **Category** | **Subcategory** | **Sub Weights** | **Weights** |
|---|---|---|---|---|
| **1** | Connectivity and technology infrastructure | Broadband penetration | 15% | 20% |
| | | Broadband quality | 10% | |
| | | Broadband affordability | 10% | |
| | | Mobile-phone penetration | 15% | |
| | | Mobile quality | 10% | |
| | | Internet user penetration | 15% | |
| | | International Internet bandwidth | 10% | |
| | | Internet security | 15% | |
| **2** | Business environment | Overall political environment | 11.1% | 15% |
| | | Macroeconomic environment | 11.1% | |
| | | Market opportunities | 11.1% | |
| | | Policy towards private enterprise | 11.1% | |
| | | Foreign investment policy | 11.1% | |
| | | Foreign trade and exchange regimes | 11.1% | |
| | | Tax regimes | 11.1% | |
| | | Financing | 11.1% | |
| | | Labor market | 11.1% | |
| **3** | Social and cultural environment | Education level | 20% | 15% |
| | | Internet literacy | 20% | |
| | | Degree of entrepreneurship | 20% | |
| | | Technical skills of workforce | 20% | |
| | | Degree of innovation | 20% | |
| **4** | Legal environment | Effectiveness of traditional legal framework | 30% | 10% |
| | | Laws covering the Internet | 25% | |
| | | Level of censorship | 10% | |
| | | Ease of registering a new business | 25% | |
| | | Electronic ID | 10% | |
| **5** | Government policy and vision | Government spend on ICT | 5% | 15% |
| | | Digital development strategy | 25% | |
| | | E-government strategy | 20% | |
| | | Online procurement | 5% | |
| | | Availability of online public services for citizens | 15% | |
| | | Availability of online public services for businesses | 15% | |
| | | E-participation | 15% | |

| No. | Category | Subcategory | Sub Weights | Weights |
|---|---|---|---|---|
| **6** | Consumer and business adoption | Consumer spending on ICT per head | 15% | 25% |
| | | Level of e-business development | 10% | |
| | | Use of Internet by consumers | 25% | |
| | | Use of online public services by citizens | 25% | |
| | | Use of online public services by businesses | 25% | |

The main objective of the present paper is to realize a classification and clustering analysis upon the digital economy ranking without using the weights for the principal characteristics used for ranking, meaning that the aggregation indicator is not applied for the clustering, but will be used for the validation of the new classification type.

## 2 Excel to arff open source convertor

In [6], the concept of open source software is defined as a free program distributed in which source code is open and visible and its main features are:

- free distribution– restriction is not permitted by license;
- source code – it should be included and open to a product distributed through open source;
- changes made on these products can be made and the resulting programs can themselves be distributed;
- the integrity of the author’s code meaning that the product’s license shows clearly whether the programs resulting from changes can be distributed with the same name as the original product or not;
- lack of discrimination – license does not discriminate any group of persons or areas in which the product is intended to be used.

Referring to the need to identify a correlation between the characteristics analyzed for the Digital Economy Ranking, and open source was needed to convert an excel file into an ARFF, format needed for the data mining analysis. This program is available at [7].

The ARFF, Attribute-Relation File Format, presented in [8], consists in the following parts:

- header information;
- data information.

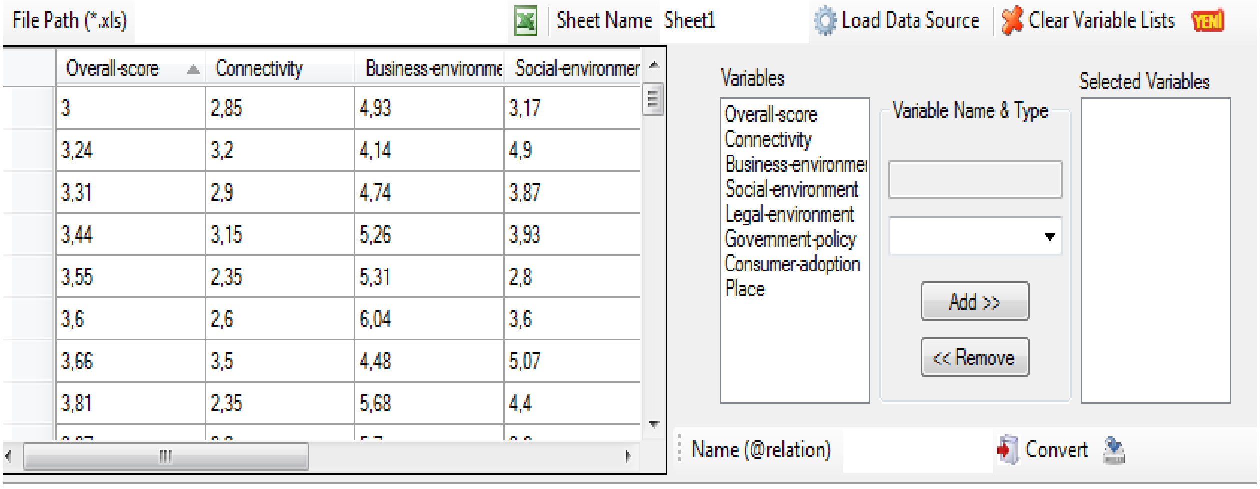
The header information part contains the name of the relation, the list of attributes, along with their type, as presented in the example below.

```
@RELATION test
@ATTRIBUTE attribute1 NUMERIC
@ATTRIBUTE attribute2 {YES, NO}
@ATTRIBUTE attribute3 DATE "yyyy-MM-dd HH:mm:ss"
```

For the data information, after the keyword @DATA, the list of data, the values accorded for each instance to the attributes described in the header part. The specific format is:

```
@DATA
13, YES, "2001-04-03 12:12:12"
12, NO, "2001-05-03 12:59:55"
```

In order to form an ARFF, using EXCELtoARFF open source, the file path must be indicated, a .xls format. Loading the Data Source, in the left side of the application, the data from the Excel specified sheet appears on the screen, as in Figure 1.



Sheet Name Sheet1

| Overall-score | Connectivity | Business-environme | Social-environmer |
|---|---|---|---|
| 3 | 2,85 | 4,93 | 3,17 |
| 3,24 | 3,2 | 4,14 | 4,9 |
| 3,31 | 2,9 | 4,74 | 3,87 |
| 3,44 | 3,15 | 5,26 | 3,93 |
| 3,55 | 2,35 | 5,31 | 2,8 |
| 3,6 | 2,6 | 6,04 | 3,6 |
| 3,66 | 3,5 | 4,48 | 5,07 |
| 3,81 | 2,35 | 5,68 | 4,4 |

Variables
Overall-score
Connectivity
Business-environme
Social-environment
Legal-environment
Government-policy
Consumer-adoption
Place

Variable Name & Type

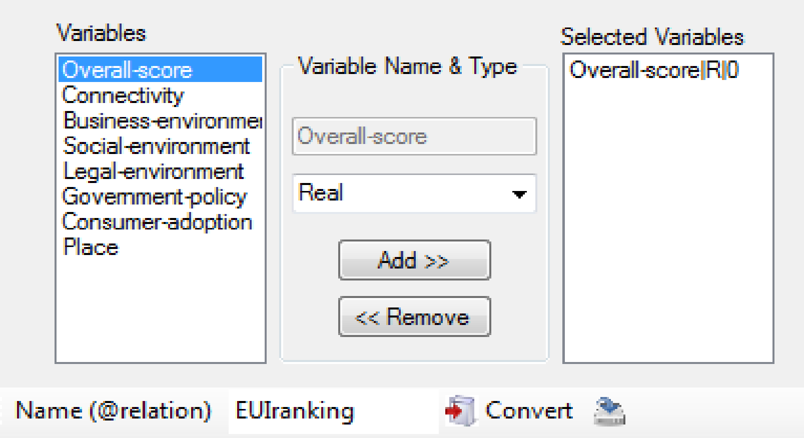
Selected Variables

Name (@relation)

**Fig. 1.** Excel data load

The next step, is to configure the heading part from the ARFF file, this leading to the right side of the framework from figure 1. The left list of variables are the variables named in the excel file, list that is automatically extracted, as each column heading name. The right list is the list of the name of attributes that will be used in the ARFF heading part. Clicking on a variable name, the type can be selected, as real or string.

In Figure 2, the variable Overall-score is selected for configuration, choosing the real type, and automatically named Overall-score|R|O, a standard prefix used, '|R|O'. Clicking on the Add button, the variables is set for the ARFF file. Each variable must be selected to be added in the ARFF file, along with its type.



Variables
Overall-score
Connectivity
Business-environme
Social-environment
Legal-environment
Government-policy
Consumer-adoption
Place

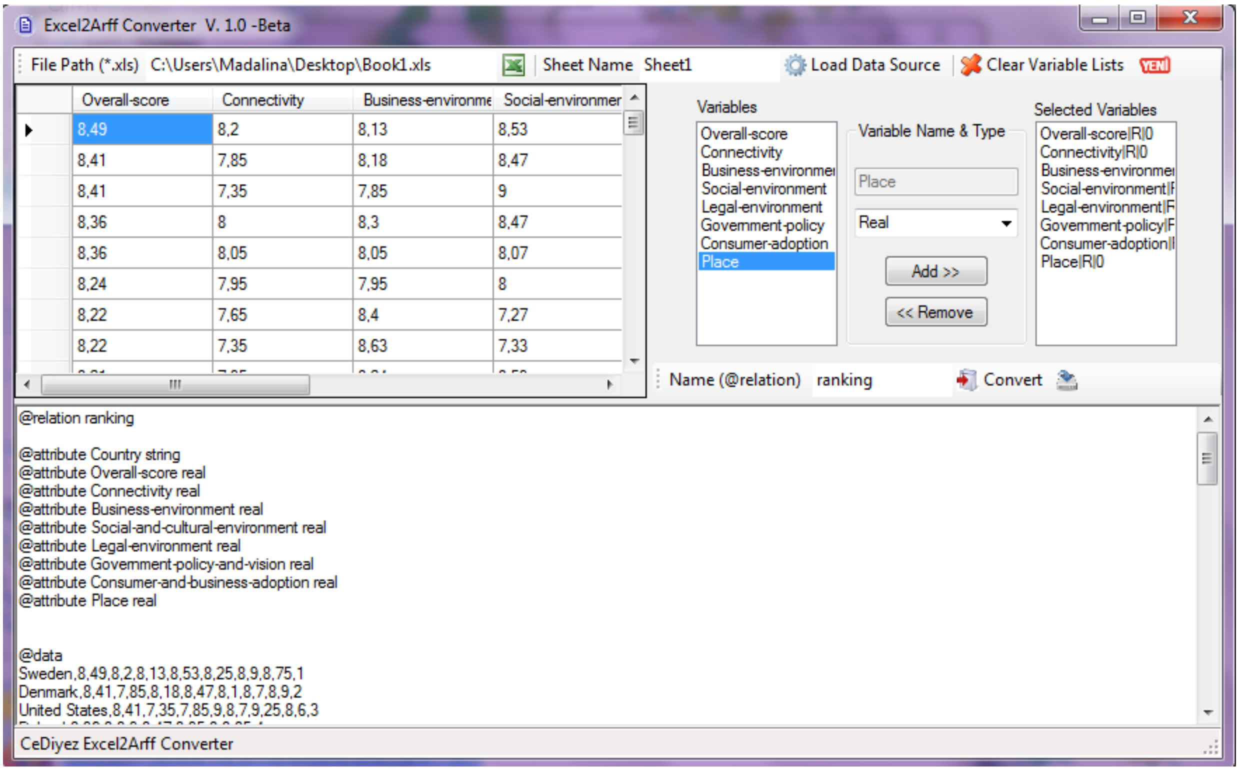
Variable Name & Type
Overall-score
Real

Selected Variables
Overall-score|R|0

Name (@relation) EUIranking

**Fig. 2.** Arff Header part configuration

Last, selecting the name of the @relation, in our case 'EUIranking', the Convert button can be clicked, generating, in the bottom part of the window, the ARFF file. In the meantime, it can be saved at a specific path, choosing the button right from the Convert one. The header and data parts generated are presented in Figure 3.



File Path (*.xls) C:\Users\Madalina\Desktop\Book1.xls

Sheet Name Sheet1

| Overall-score | Connectivity | Business-environme | Social-environmer |
|---|---|---|---|
| 8,49 | 8,2 | 8,13 | 8,53 |
| 8,41 | 7,85 | 8,18 | 8,47 |
| 8,41 | 7,35 | 7,85 | 9 |
| 8,36 | 8 | 8,3 | 8,47 |
| 8,36 | 8,05 | 8,05 | 8,07 |
| 8,24 | 7,95 | 7,95 | 8 |
| 8,22 | 7,65 | 8,4 | 7,27 |
| 8,22 | 7,35 | 8,63 | 7,33 |

Variables:
Overall-score
Connectivity
Business-environmei
Social-environment
Legal-environment
Government-policy
Consumer-adoption
Place

Variable Name & Type: Place, Real

Selected Variables:
Overall-score|R|0
Connectivity|R|0
Business-environmei
Social-environment|F
Legal-environment|F
Government-policy|F
Consumer-adoption|I
Place|R|0

Name (@relation) ranking

```
@relation ranking

@attribute Country string
@attribute Overall-score real
@attribute Connectivity real
@attribute Business-environment real
@attribute Social-and-cultural-environment real
@attribute Legal-environment real
@attribute Government-policy-and-vision real
@attribute Consumer-and-business-adoption real
@attribute Place real


@data
Sweden,8,49,8,2,8,13,8,53,8,25,8,9,8,75,1
Denmark,8,41,7,85,8,18,8,47,8,1,8,7,8,9,2
United States,8,41,7,35,7,85,9,8,7,9,25,8,6,3
```

CeDiyez Excel2Arff Converter

**Fig. 3.** Arff file Convert

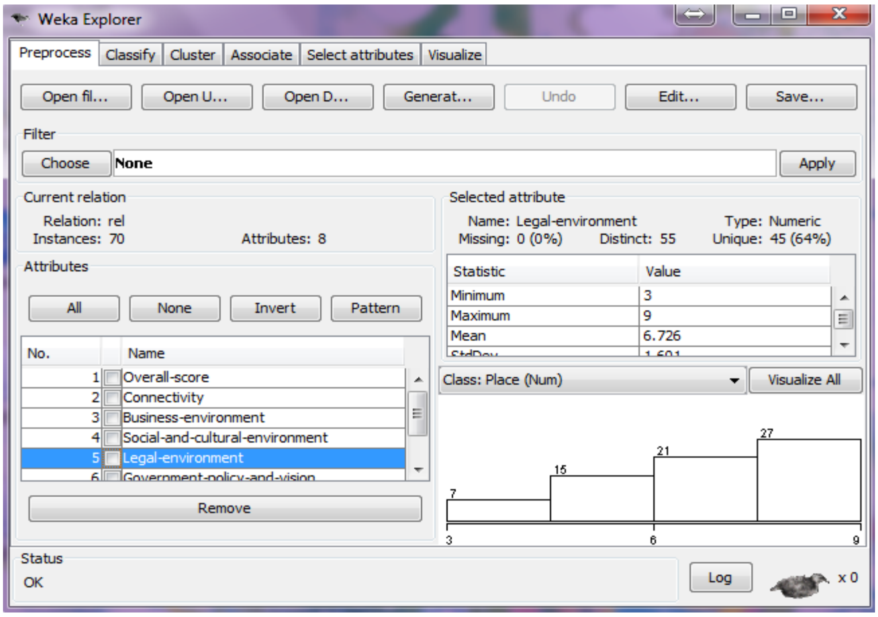
This conversion is done with no loose of information. The difference between the excel file and the arff one, is the one given by the specific header that the ARFF needs in order to determine the meaning of the data that follows it. The new file format generated, the next step, the one of data mining process, is done using another open source tool, Weka Data Mining, tool presented in the next chapter.

## 3 Weka Data Mining open source tool

Weka, Waikato Environment for Knowledge Analysis, is a suite of machine learning software, developed by University of Waikato, New Zeeland, with the website [10]. The available operations that can be achieved in Weka open source tool are:

- preprocessing phase;
- data classification;
- data clustering;
- association rules;
- attribute selection;
- data visualization.

In the present section, all 6 phases available within Weka software are illustrated with the data generated by ExceltoArff convertor, the ARFF file containing six characteristics for each 70 countries analyzed. For the first step, the preprocess one, data must be indicated, meaning the ARFF file containing the relation, attributes description and data information. Figure 4 reveals the preprocess step, after inserting the path file for Digital Economy Ranking data file.



Filter

None

Current relation

Relation: rel
Instances: 70 Attributes: 8

Selected attribute

Name: Legal-environment Type: Numeric
Missing: 0 (0%) Distinct: 55 Unique: 45 (64%)

| Statistic | Value |
|---|---|
| Minimum | 3 |
| Maximum | 9 |
| Mean | 6.726 |
| StdDev | 1.601 |

Attributes

| No. | Name |
|---|---|
| 1 | Overall-score |
| 2 | Connectivity |
| 3 | Business-environment |
| 4 | Social-and-cultural-environment |
| 5 | Legal-environment |
| 6 | Government-policy-and-vision |

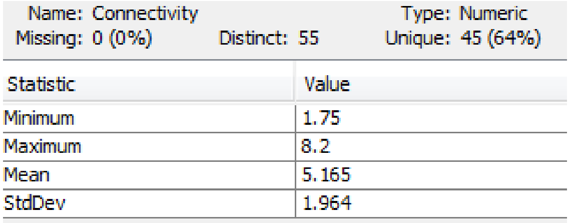
Class: Place (Num)

Status

OK

**Fig. 4.** Preprocessing phase

For a deeper view, Figure 5 contains the statistic indicators calculated for each characteristic. The minimum value for connectivity is situated for Nigeria, 62th place, while the maximum value of 8.2 is pointed for Sweden, the country situated in the first place. The mean value of the 70 countries is 5.1, with a standard deviation of 1.964.



Name: Connectivity Type: Numeric
Missing: 0 (0%) Distinct: 55 Unique: 45 (64%)

| Statistic | Value |
|---|---|
| Minimum | 1.75 |
| Maximum | 8.2 |
| Mean | 5.165 |
| StdDev | 1.964 |

**Fig. 5.** Statistic information for connectivity characteristic

Other information available in figure 4 is the one relating the attribute type, name, as:

- name, the name of the attribute selected;
- type, the attribute type, that can be real, string;
- missing, the percentage of missing values of the specific attribute;
- distinct, the number of distinct values that attribute analyzed has;
- unique, the number and percentage of instances in the data having a value for this attribute that no other instances have.

The filter section allows working with filters upon the data, already created filters, or ones implemented by the users. The types of filters available are:

- supervised;
- unsupervised.

The supervised filters are also divided into:

- attribute filters, such as add classification, attribute selection, nominal to binary;
- instance filters, like stratified remove folds.

For the unsupervised filters, attribute and instances are available to. For the attribute type are present:

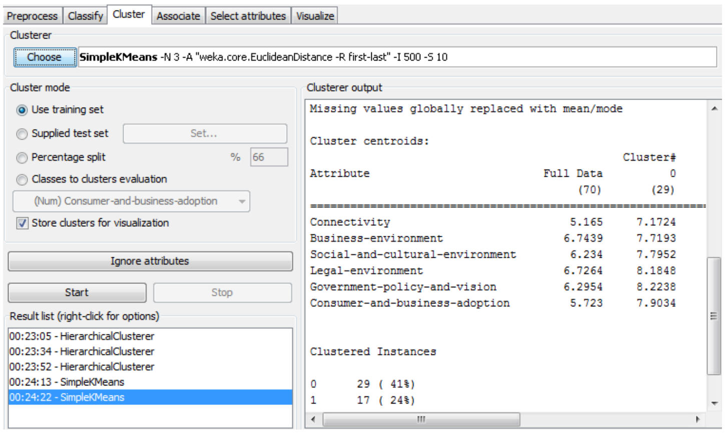
- add, add expression, add id, add noise, add values;
- center;
- copy;
- discrete;
- nominal to binary;

- nominal to string;
- kernel;
- reorder;
- standardize.

When it comes to unsupervised filters for the instances, the following types occur:

- remove folds;
- normalize;
- randomize;
- resample;
- remove range;
- remove percentage.

The filters being applied, we move to the clustering zone, presented in Figure 6.



SimpleKMeans -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

● Use training set
○ Supplied test set
○ Percentage split % 66
○ Classes to clusters evaluation
(Num) Consumer-and-business-adoption
☑ Store clusters for visualization

Result list (right-click for options)

00:23:05 - HierarchicalClusterer
00:23:34 - HierarchicalClusterer
00:23:52 - HierarchicalClusterer
00:24:13 - SimpleKMeans
00:24:22 - SimpleKMeans

```
Missing values globally replaced with mean/mode

Cluster centroids:
                                                    Cluster#
Attribute                              Full Data           0
                                            (70)        (29)
==============================================================
Connectivity                               5.165      7.1724
Business-environment                      6.7439      7.7193
Social-and-cultural-environment            6.234      7.7952
Legal-environment                         6.7264      8.1848
Government-policy-and-vision              6.2954      8.2238
Consumer-and-business-adoption             5.723      7.9034


Clustered Instances

0      29 ( 41%)
1      17 ( 24%)
```

**Fig. 6.** Clustering analyses in Weka

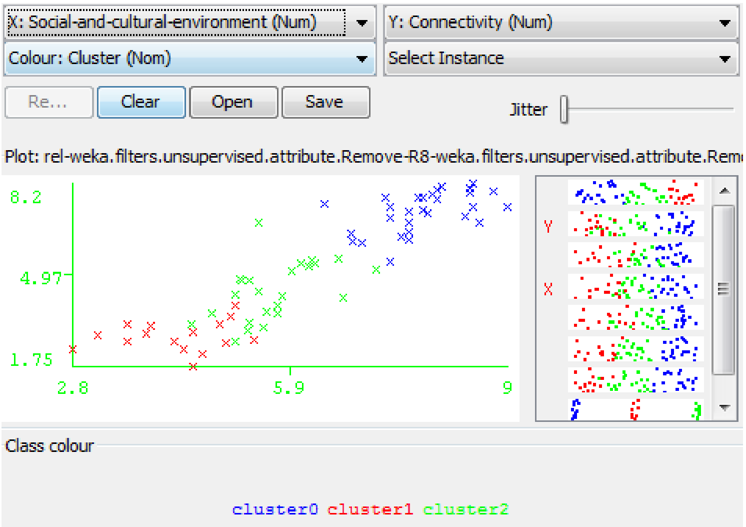
First, a cluster type is chosen, from the available list:

- Farthest First;
- Filtered Clustered;
- Hierarchical;
- Simple k Means;
- X Means.

A selection of attributes ignoring can be applied. The output of the clustering contains:

- the number of iterations needed for the clustering process;
- the sum of squared errors within cluster;
- clusters centroids;
- percentage of instances associated to each cluster.

For each relation between the characteristics available, a 2D graphic is formed, for the X axis a characteristic, and for the Y axis another chosen characteristic. Each point from this XOY space has different colors, indicating the membership to its cluster, as in Figure 7.

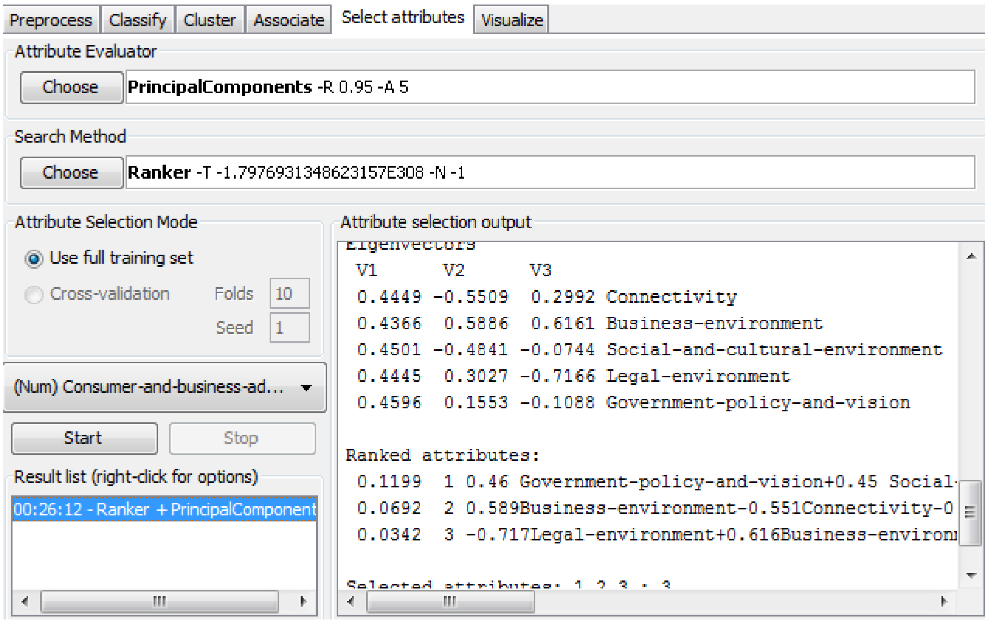


X: Social-and-cultural-environment (Num)
Y: Connectivity (Num)
Colour: Cluster (Nom)
Select Instance

Plot: rel-weka.filters.unsupervised.attribute.Remove-R8-weka.filters.unsupervised.attribute.Rem

**Fig. 7.** Two attribute clustering membership

Attribute selection Weka components deals with principal components analysis, analysis done in order to minimize the redundancy among the information within the instances from the relation. For the example of Digital Economy Ranking, the six characteristics saved are transformed with 3 new vectors used for information explanation, ranked attributes. A transformation expression is generated for attribute to eigenvectors conversion, with as less as possible loose of information, but with a lowering in the total dimension of data. In Figure 8, the eigenvectors generated are available.



**PrincipalComponents** -R 0.95 -A 5

**Ranker** -T -1.7976931348623157E308 -N -1

● Use full training set

○ Cross-validation Folds 10 Seed 1

(Num) Consumer-and-business-ad...

00:26:12 - Ranker + PrincipalComponent

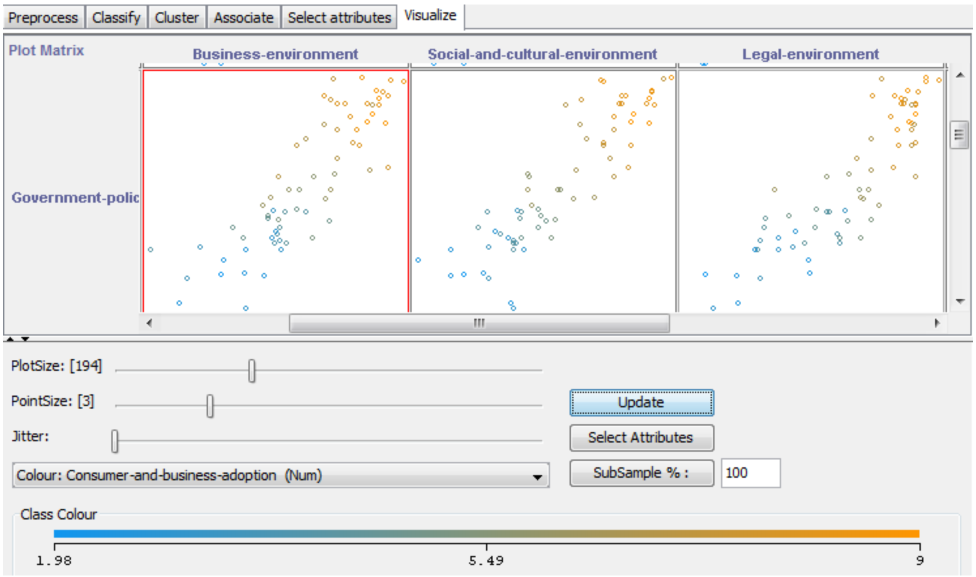
```
Eigenvectors
 V1       V2       V3
 0.4449 -0.5509  0.2992 Connectivity
 0.4366  0.5886  0.6161 Business-environment
 0.4501 -0.4841 -0.0744 Social-and-cultural-environment
 0.4445  0.3027 -0.7166 Legal-environment
 0.4596  0.1553 -0.1088 Government-policy-and-vision

Ranked attributes:
 0.1199  1 0.46 Government-policy-and-vision+0.45 Social
 0.0692  2 0.589Business-environment-0.551Connectivity-0
 0.0342  3 -0.717Legal-environment+0.616Business-environ

Selected attributes: 1,2,3 : 3
```

**Fig. 8.** Weka principal components stage

As for the last phase, the visualization offers a 2D spatial representation of the interaction between the attributes analyzed, Figure 9.



PlotSize: [194]

PointSize: [3]

Jitter:

Colour: Consumer-and-business-adoption (Num)

SubSample % : 100

**Fig. 9.** Weka visualization stage

The advantages given by Weka open source include, as seen in [9]:

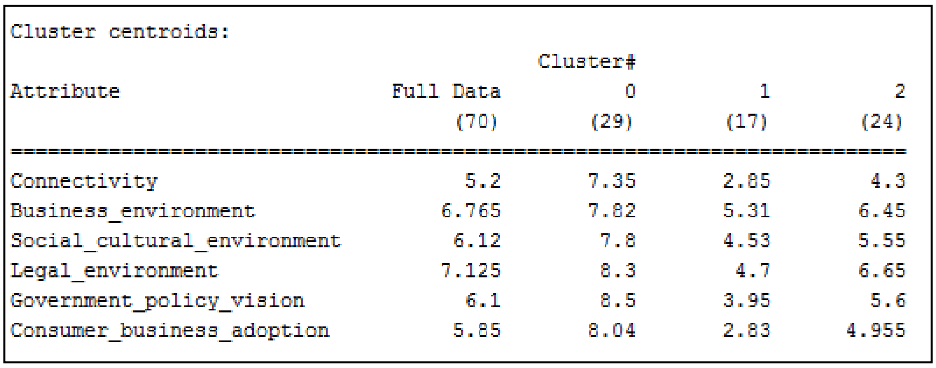
- freely available under GNU General Public License;
- portability;

- a collection of data preprocessing and modeling techniques;
- graphical user interfaces.

## 4 Implementation results

For an in-depth documentary concerning the suitable spatial models used in the process of classification, a brief overview of the main methods used in literature is done. Clustering analysis is the method used in data mining, information retrieval, pattern recognition and is a spatial representation model that is defined as an assignment of a set of objects into smaller subsets, called clusters, by the similarity between the objects from the same cluster and the differences among the clusters.

Applying the k-means for the 70 registrations with the 6 characteristics, with a number of three clusters, the following centroids resulted from Figure 10. Cluster 1 is associated to the first 24 countries, cluster 2 is for the next 17 countries, while cluster 0 is formed out of the last 29 countries.



```
Cluster centroids:
                                               Cluster#
Attribute                          Full Data          0          1          2
                                        (70)       (29)       (17)       (24)
=============================================================================
Connectivity                             5.2       7.35       2.85        4.3
Business_environment                   6.765       7.82       5.31       6.45
Social_cultural_environment             6.12        7.8       4.53       5.55
Legal_environment                      7.125        8.3        4.7       6.65
Government_policy_vision                 6.1        8.5       3.95        5.6
Consumer_business_adoption              5.85       8.04       2.83      4.955
```

**Fig. 10.** Cluster's centroids for k-means algorithm

For a correlation example between two categories used for the present ranking, business environment, OX axis, and social and cultural environment, OY axis, is formed in the graphic from Figure 11.

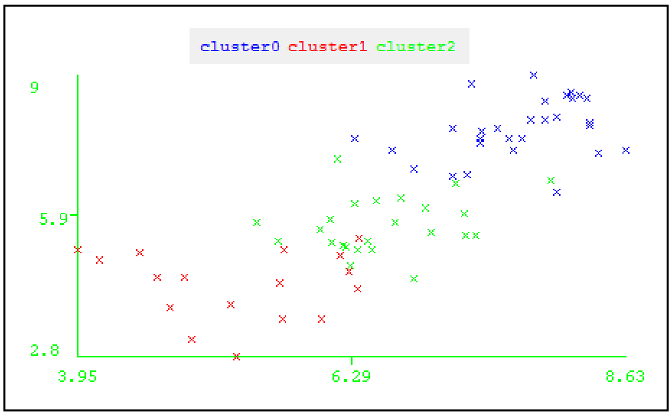
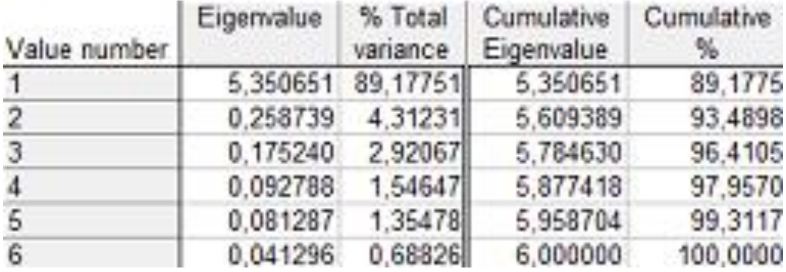


**Fig. 11.** Two characteristics clustering correlation

The following analyze is done using first the method of principal components extraction, called eigenvalues, from the six classification available. Table 2 contains the cumulative percentage of information gathered in the characteristics.

**Table 1.** Eigenvalues



| Value number | Eigenvalue | % Total variance | Cumulative Eigenvalue | Cumulative % |
|---|---|---|---|---|
| 1 | 5,350651 | 89,17751 | 5,350651 | 89,1775 |
| 2 | 0,258739 | 4,31231 | 5,609389 | 93,4898 |
| 3 | 0,175240 | 2,92067 | 5,784630 | 96,4105 |
| 4 | 0,092788 | 1,54647 | 5,877418 | 97,9570 |
| 5 | 0,081287 | 1,35478 | 5,958704 | 99,3117 |
| 6 | 0,041296 | 0,68826 | 6,000000 | 100,0000 |

Selecting the first two eigenvalues, the next 2D diagram includes the 70 countries and total information of 93.48% from the total one obtained within the six characteristics analyzed.

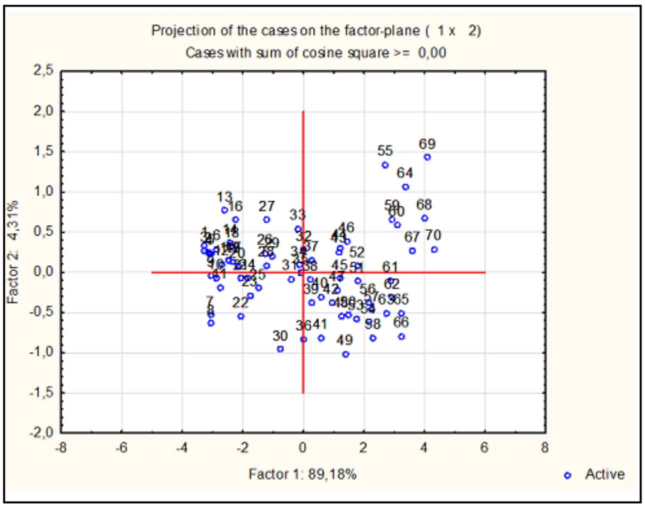
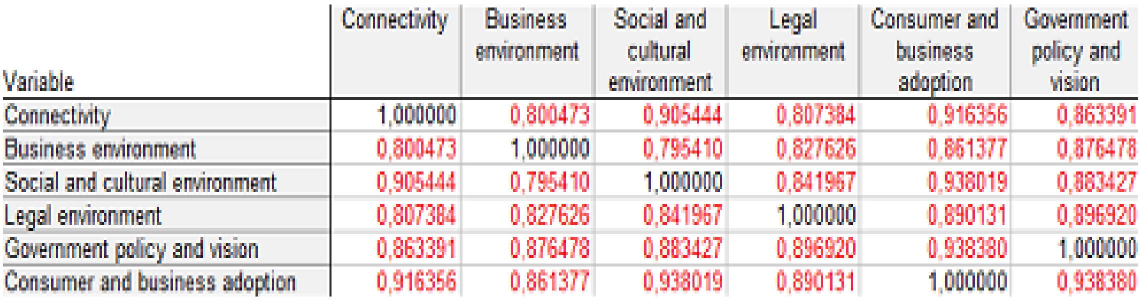


**Fig. 12.** Two eigenvalues graphic

It can be seen that, if only the first 2 new dimensions are selected, a percentage of 93 from the total information is condensed in a 2D plan. The correlation within the six characteristics is revealed in Table 3.

**Table 2.** Correlation matrix



| Variable | Connectivity | Business environment | Social and cultural environment | Legal environment | Consumer and business adoption | Government policy and vision |
|---|---|---|---|---|---|---|
| Connectivity | 1,000000 | 0,800473 | 0,905444 | 0,807384 | 0,916356 | 0,863391 |
| Business environment | 0,800473 | 1,000000 | 0,795410 | 0,827626 | 0,861377 | 0,876478 |
| Social and cultural environment | 0,905444 | 0,795410 | 1,000000 | 0,841967 | 0,938019 | 0,883427 |
| Legal environment | 0,807384 | 0,827626 | 0,841967 | 1,000000 | 0,890131 | 0,896920 |
| Government policy and vision | 0,863391 | 0,876478 | 0,883427 | 0,896920 | 0,938380 | 1,000000 |
| Consumer and business adoption | 0,916356 | 0,861377 | 0,938019 | 0,890131 | 1,000000 | 0,938380 |

For the most correlated elements, the consumer and business adoption and social and cultural environment, a regression is formed in Figure 13.

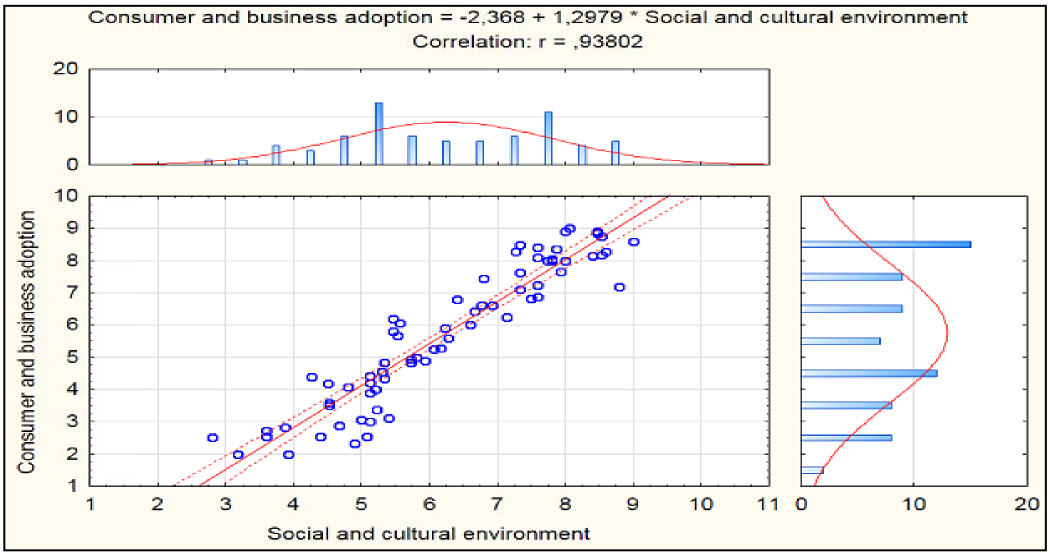


**Fig. 13.** Regression between consumer and social cultural environment

In Figure 14, the six indicators are clustered referring to the distance between them. It can be seen that business and legal environment, and also social and cultural environment with government policy and vision, are two sets of indicators that can be clustered first to determine an aggregated value.

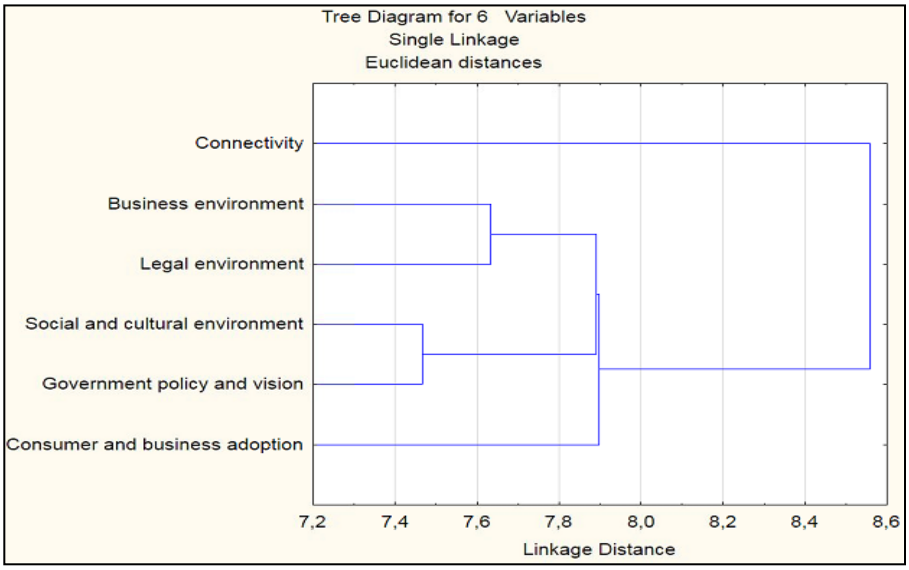


**Fig. 14.** Characteristics clustering

Using clustering analysis with City Block distances, along with correlation and regression forming, the results reveal a powerful correlation impact between all the six characteristics recorded.

## 5 Conclusions

The main objective of the digital economy ranking is the one of analyzing the stage of the impact of using the digital tools in the central economy of each country. Referring to the present analysis, the use of principal components extraction offers a redundancy lowering in the information used for the ranking of the level of e-readiness of each country.

Conversion open source tools are required in order to minimize the costs of transformation between different formats needed for each software product. The present paper demonstrated the need of this open source, because of the initial data format, an .xls file, that had to be turned into an ARFF file, the standard input for Weka Data Mining software.

Removing the general weights used for information aggregation didn't generate a modification of ranking position in the classification. Spatial representation, along with multidimensional analysis such as clustering, offers a proper interpretation of the comparison between the 70 countries available.

Clustering using k-means algorithm naturally transform one cluster into a number of multiple clusters, with the particularity of minimizing the distances between the objects from the same cluster and maximizing the distances between the objects of different clusters. This method offers a dense representation and interpretation. For a proper clustering, knowing the centroid of it is enough to characterize the whole cluster.

## Acknowledgments

This work was cofinanced from the European Social Fund through Sectorial Operational Programme Human Resources Development 2007-2013, project number POSDRU/107/1.5/S/77213 „Ph.D. for a career in interdisciplinary economic research at the European standards".

## References

[1] E-readiness defition, Available online at http://en.wikipedia.org/wiki/E-readiness

[2] E. Brynjolfsson and B. Kahin – "Understanding the Digital Economy; Data, Tools and Research", *MIT Press*, 2002, 372 pages, ISBN 0-262-02474-8

[3] V. Maugis, N. Choucri, S. Madnick, M. Siegel, S. Gillett, F. Haghseta, H. Zhu and M. Best – "Global e-Readiness – For What? Readiness for e-Banking(JITD)", *MIT Sloan School of Management*, 2004, 33 pp

[4] B. F. Schmid – "What is new about the Digital Economy?", *Electronic Markets*, Vol. 11, No. 1, 2001, pp. 44-51, ISSN 1422-8890

[5] Economist Intelligence Unit – "Digital economy ranking 2010. Beyond e-readiness", Available online at http://www.eiu.com/sso/cas/client

[6] M. Doinea, "Open Source Security – Quality Requests", Open Source Science Journal, Vol. 1, No. 1, 2009, pg. 126-135, ISSN 2066 – 740X

[7] Excel to Arff Converter, Available online at http://sourceforge.net/projects/exceltoarffconv/

[8] Attribute-Relation File Format, Available online at http://www.cs.waikato.ac.nz/~ml/weka/arff.html

[9] Weka definition, http://en.wikipedia.org/wiki/Weka_(machine_learning)

[10] Weka, Available online at http://www.cs.waikato.ac.nz/ml/weka/

[11] Weka Documentation, Available online at http://netcologne.dl.sourceforge.net/project/weka/documentation/3.5.x/ExperimenterTutorial-3-5-8.pdf

**Mădălina ZURINI** is currently a PhD candidate in the field of Economic Informatics. She graduated the Faculty of Cybernetics, Statistics and Economic Informatics (2008) and a master in Computer Science, having her dissertation given in *Implications of Bayesian classifications for optimizing spam filters* (2010). She is also engaged in Pedagogical Program as part of the Department of Pedagogical Studies. Her fields of interest are data classification, artificial intelligence, data quality, algorithm analysis and optimizations. She wants to pursue a pedagogical career.