| Economic Insights – Trends and Challenges | Vol.III (LXVI) No. 2/2014 | 33 - 40 |
|---|---|---|

# Clustering HRST Time Series of the EU Countries

Cristian Marinoiu

Petroleum-Gas University of Ploieşti, Bd. Bucureşti 39, 100680, Ploieşti, Romania
e-mail: marinoiu_c@yahoo.com

## Abstract

*HRST (human resource in science and technology) is one of the indicators which reflects a country's degree of implication in supporting the development of the science and technology field as an important factor of the economic and social progress. In this paper we propose a EU classification based on similarities in the evolution of HRST, during the period 2002-2012. The methodology involves clustering time series.*

**Key words:** *clustering time series, human resource in science and technology*

**JEL Classification:** *C38, J21*

## Introduction

In 2010 the European Commission launched the strategy Agenda 2020. The priorities established in the Agenda are [2]:

- The smart economic growth through more efficient investments in education, research and innovation;
- The sustainable economic growth based on the decrease of carbon dioxide emissions;
- The economic growth favorable to social inclusion through creating new jobs and through poverty reduction;

In order to implement the announced priorities, the European Commission established objectives to be achieved until 2020 in the following fields: employment, research and development, climate changes and sustainable use of the energy, education and fight against poverty and social exclusion. The research and the innovation in science and technology represent important factors of the economic and social progress. The evolution of this field in each EU country is reflected by the values of the specific indicators provided by Eurostat [3]. HRST is defined as the percentage of persons aged between 25 and 64 who work in science and technology, relative to the total labour force. In this paper we propose a classification of EU countries based on the similarity of the evolution of this indicator during the period 2002-2012.

## HRST Indicator

The annual values of the HRST indicator measured during the period 2002-2012 for each EU country [3] are presented in table 1.

**Table 1.** The annual values of HRST indicator during the period 2002-2012

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Belgium | 42.9 | 43.7 | 44.9 | 46.2 | 46.6 | 46.7 | 47.0 | 48.2 | 49.3 | 49.6 | 50.3 |
| Bulgaria | 31.2 | 31.6 | 31.2 | 31.6 | 30.5 | 30.8 | 31.0 | 32.2 | 31.6 | 33.0 | 32.6 |
| Czech Republic | 31.6 | 32.3 | 32.8 | 34.5 | 34.8 | 36.0 | 37.1 | 37.9 | 37.8 | 36.0 | 36.5 |
| Denmark | 45.6 | 46.9 | 46.9 | 49.1 | 50.4 | 48.8 | 49.4 | 50.0 | 51.0 | 51.5 | 52.9 |
| Germany | 41.5 | 42.2 | 42.7 | 43.1 | 43.2 | 43.6 | 44.0 | 44.7 | 44.8 | 44.9 | 45.7 |
| Estonia | 40.0 | 40.0 | 41.5 | 44.8 | 44.1 | 44.4 | 44.2 | 45.6 | 45.0 | 47.0 | 48.8 |
| Ireland | 35.6 | 37.7 | 39.2 | 39.1 | 39.5 | 41.2 | 42.2 | 44.5 | 46.0 | 49.0 | 50.5 |
| Greece | 26.2 | 27.1 | 29.4 | 29.3 | 30.8 | 31.2 | 31.7 | 31.8 | 32.4 | 33.6 | 34.2 |
| Spain | 35.0 | 35.2 | 36.6 | 38.6 | 39.8 | 39.7 | 39.7 | 39.0 | 39.0 | 40.4 | 40.6 |
| France | 37.1 | 38.5 | 39.1 | 40.2 | 41.2 | 41.7 | 42.6 | 43.3 | 43.8 | 48.1 | 48.1 |
| Croatia | 27.6 | 27.6 | 27.9 | 28.2 | 29.2 | 28.8 | 29.9 | 31.5 | 32.1 | 30.9 | 32.3 |
| Italy | 30.3 | 30.7 | 32.5 | 32.8 | 34.6 | 35.6 | 35.3 | 34.3 | 33.8 | 34.4 | 34.4 |
| Cyprus | 39.7 | 40.7 | 39.7 | 38.8 | 40.2 | 42.5 | 43.7 | 43.0 | 44.0 | 47.1 | 48.5 |
| Latvia | 33.5 | 31.6 | 31.0 | 32.7 | 34.8 | 37.2 | 39.9 | 38.9 | 37.8 | 38.2 | 40.0 |
| Lithuania | 32.3 | 32.9 | 34.6 | 37.4 | 38.3 | 40.6 | 42.5 | 41.7 | 42.7 | 43.7 | 44.2 |
| Luxembourg | 36.3 | 35.9 | 43.4 | 43.4 | 43.0 | 43.3 | 45.5 | 55.3 | 55.9 | 57.1 | 58.6 |
| Hungary | 29.0 | 30.2 | 31.8 | 31.6 | 31.9 | 31.7 | 33.2 | 33.2 | 33.0 | 34.6 | 35.4 |
| Malta | 25.9 | 27.4 | 28.4 | 29.9 | 30.4 | 31.9 | 32.1 | 32.3 | 32.1 | 34.9 | 36.4 |
| Netherlands | 45.8 | 48.2 | 49.4 | 49.3 | 48.1 | 49.8 | 50.5 | 50.9 | 51.9 | 52.2 | 52.2 |
| Austria | 33.4 | 32.8 | 40.7 | 37.9 | 38.3 | 37.6 | 37.8 | 39.0 | 39.2 | 40.5 | 41.9 |
| Poland | 25.6 | 27.4 | 28.3 | 29.6 | 31.4 | 32.5 | 33.4 | 34.9 | 36.3 | 37.0 | 37.7 |
| Portugal | 17.6 | 18.2 | 21.2 | 21.5 | 22.0 | 22.1 | 23.1 | 23.5 | 23.9 | 27.0 | 28.7 |
| Romania | 20.8 | 20.5 | 21.2 | 22.0 | 22.8 | 23.0 | 23.8 | 24.1 | 24.4 | 25.8 | 25.7 |
| Slovenia | 32.3 | 34.9 | 35.8 | 37.3 | 38.8 | 38.9 | 40.1 | 40.6 | 40.8 | 42.4 | 42.8 |
| Slovakia | 28.5 | 29.0 | 28.8 | 30.7 | 31.6 | 31.8 | 32.0 | 32.0 | 33.5 | 34.1 | 32.5 |
| Finland | 45.5 | 45.5 | 47.3 | 48.0 | 48.7 | 49.6 | 50.1 | 50.7 | 50.6 | 52.6 | 53.7 |
| Sweden | 44.7 | 45.6 | 46.3 | 47.3 | 48.0 | 48.7 | 49.3 | 49.7 | 50.3 | 51.7 | 52.6 |
| United Kingdom | 38.0 | 39.2 | 40.7 | 41.2 | 42.5 | 43.3 | 42.7 | 44.4 | 45.1 | 52.0 | 53.1 |

*Source:* Eurostat [3]

Each line of the Table 1 represents the values of a time series which characterizes the evolution of HRST indicator of the respective country. The graphic representation of the 28 time series is presented in Figure 1.

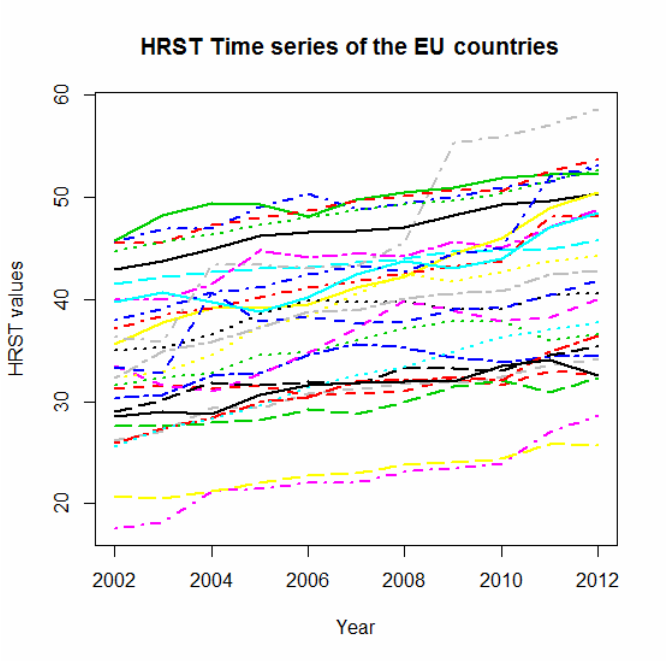


**Fig. 1.** The time series HRST of EU countries

Source: made by the author in R using data from Table 1

As a first observation we note the general upward trend of all time series, which shows the concern of EU countries in order to sustain the research and development fields during the reporting period.

## Clustering HRST Time Series

The cluster analysis aimed at partitioning a set of objects in groups named clusters so that the similarity between the objects of a cluster be minimized and the dissimilarity between the objects of different clusters be maximized. The degree of similarity between objects is established on the basis of the value of the distance between objects, distance which can be defined in various ways. Of the most frequently used options we mention the Euclidean, Mahalanobis, Minkovsky distances etc. The purpose of the cluster analysis in the case of time series is to group time series whose evolution in time is similar in a cluster. The dynamic structure of time series as well as the autocorrelation of their values make inopportune the interpretation of the time series as being only simple points in a multidimensional space. For this reason, in order to establish the degree of similarity between two time series, some special distances were proposed, including [4]: Frechet distance, dynamic time warping (DTW) distance, autocorrelation-based distance etc.

In order to detect the structure of the HRST time series we used the hierarchical agglomerative clustering [1] based on unweighted pair group method average (UPGMA), as a distance between two clusters and DTW distance as a distance between two objects. The dendogram obtained is presented in Figure 2.

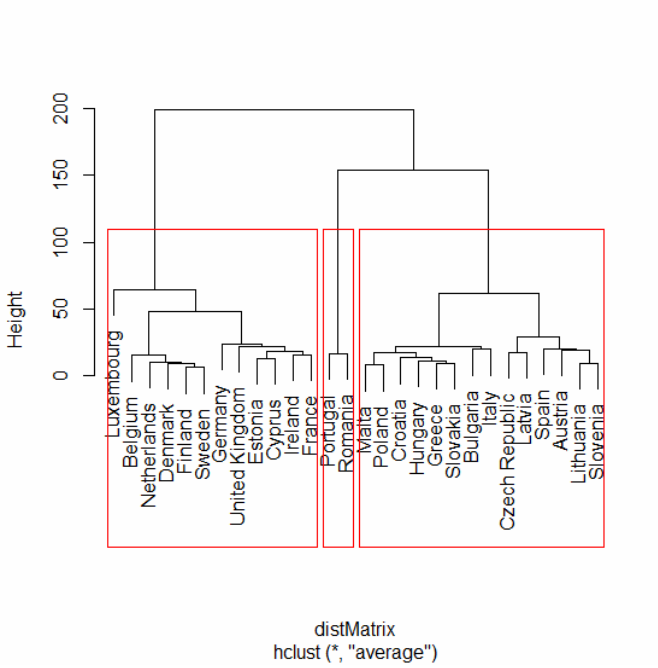


**Fig. 2.** Dendogram from agglomerative hierarchical clustering
Source: made by the author in R, using data from Table 1

The dendogram emphasizes a large number of clusters. Taking into account the fact that the height of each node in the dendogram is directly proportional to the distance between left and right sub-branch cluster, we cut the branches of the *dendogram* at height h=100, obtaining three highlighted clusters:

Cluster 1: Belgium, Denmark, Germany, Estonia, Ireland, France, Cyprus, Luxembourg Netherlands, Finland, Sweden, United Kingdom;

Cluster 2: Bulgaria, Czech Republic, Greece, Spain, Croatia, Italy, Latvia, Lithuania, Hungary, Malta, Austria, Poland, Slovenia, Slovakia;

Cluster 3: Portugal, Romania.

## The Validation of the Obtained Structure

For validating the number of clusters as well as the obtained structure we calculated and represented the silhouette of each object (time series). The silhouette of an object is a measure of the degree of affiliation of that object to the cluster in which it is classified. In accordance with [5], the values of the silhouette of an object belong to the interval [-1,1] and have the following interpretation:

- $s(i) \approx 1$, object $i$ is well classified ;
- $s(i) \approx 0$, object $i$ lies intermediate between the cluster in which it was classified and the nearest cluster;
- $s(i) \approx -1$, object $i$ is badly classified,

An indicator of the quality of a structure of clusters is given by the overall average silhouette width; the higher this value the better the quality of the structure. The dendogram in figure 2 suggests that the structures of interest could be the ones obtained for a number of clusters k equal to 2, 3, 4, 5 and, respectively, 6. The graphic representation in figure 3 of the values

overall average silhouette width versus the number of clusters shows that the maximum of these values, named Silhouette Coefficient (SC) is 0.67 and it is achieved for k=3. In figure 3 we represent the overall average silhouette width versus the number of clusters. It can be noticed that the maximum of these values, named Silhouette Coefficient (SC) is 0.67 and it is achieved for k=3. According to [5], page 10, the interpretation of the values of the Silhouette Coefficient is as follows:

- if $0.71 \leq SC \leq 1.00$ a strong structure has been found;
- if $0.51 \leq SC \leq 0.70$ a reasonable structure has been found;
- if $0.26 \leq SC \leq 0.50$ the structure is weak and could be artificial;
- if $SC \leq 0.25$ no substantial structure has been found;

Thus, we can say, that in our case, a reasonable structure has been found, for k=3.

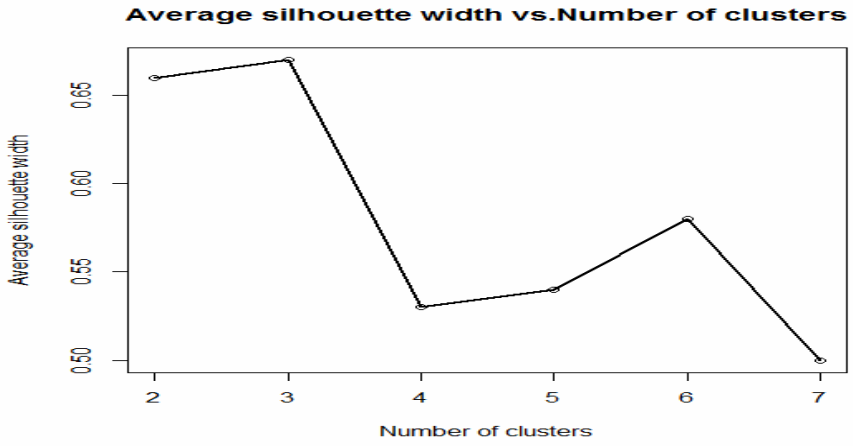


**Fig. 3.** Graph of overall average silhouette width versus number of clusters
Source: made by the author in R

Also, by analyzing the graphs of the silhouettes in figure 4, we observe that the average silhouette widths are: 0.75 for cluster 1, 0.57 for cluster 2 and 0.89 for cluster 3, values which confirm once more the good quality of the obtained structure.

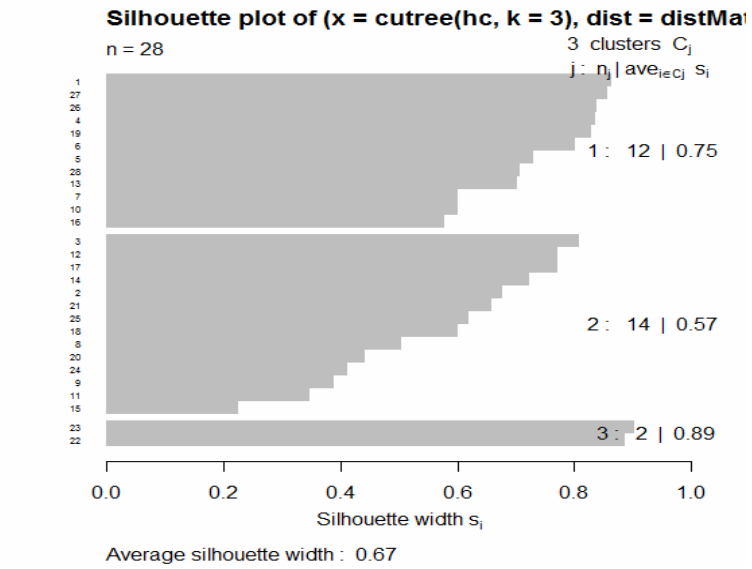


**Fig. 4.** Silhouette plot of the obtained structure

Source: made by the author in R

## Results and Discussion

In the following, we will present some considerations about the obtained results. Thus, we notice the similar evolutions of HRST indicator for most developed countries from the Western and Northern Europe into Cluster 1 (Belgium, Denmark, Germany, Estonia, Ireland, France, Luxembourg, Netherlands, Finland, Sweden, United Kingdom), the notable exceptions being here the former socialist countries Estonia and Cyprus (see figure 5) .

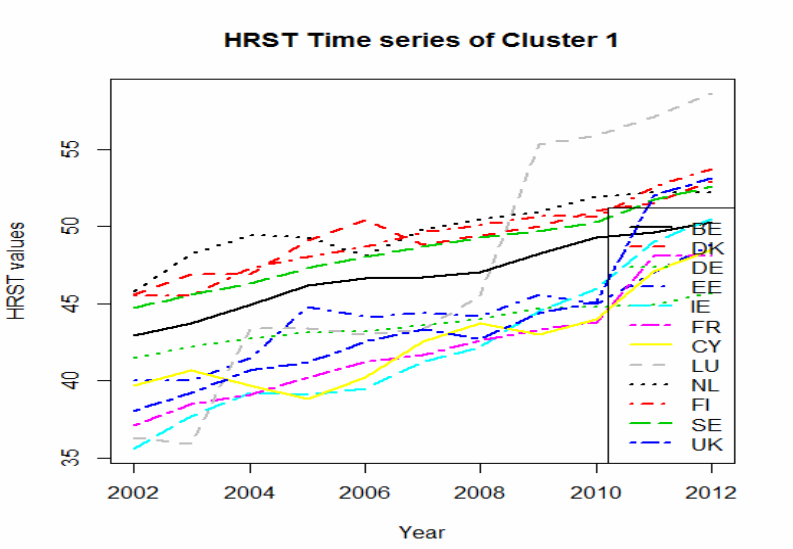


**Fig. 5.** HRST Time series of Cluster 1 *Source:* made by the author in R

In Table 2, we present, for each country from Cluster 1, the means and standard deviations of the HRST indicator values.

**Table 2.** Means and standard deviations of HRST indicator values for the countries from Cluster 1

| Country[1] | BE | DK | DE | EE | IE | FR | CY | LU | NL | FI | SE | UK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 46.8 | 49.3 | 43.7 | 44.1 | 42.2 | 42.1 | 42.5 | 47.1 | 49.8 | 49.3 | 48.6 | 43.8 |
| Std Dev | 2.4 | 2.2 | 1.3 | 2.7 | 4.7 | 3.6 | 3.1 | 8.2 | 1.1 | 2.6 | 2.5 | 4.8 |

Source: made by the author in R

The Cluster 2 is dominated by the former socialist countries (Bulgaria, Czech Republic, Croatia, Latvia, Lithuania, Hungary, Poland, Slovenia, Slovakia) along with countries like Greece, Spain, Italy, Malta, Austria (see Figure 6).

[1]Europe ISO country code (ISO-3166-2) available at http://www.countrycallingcodes.com/iso-country-codes/europe-codes.php

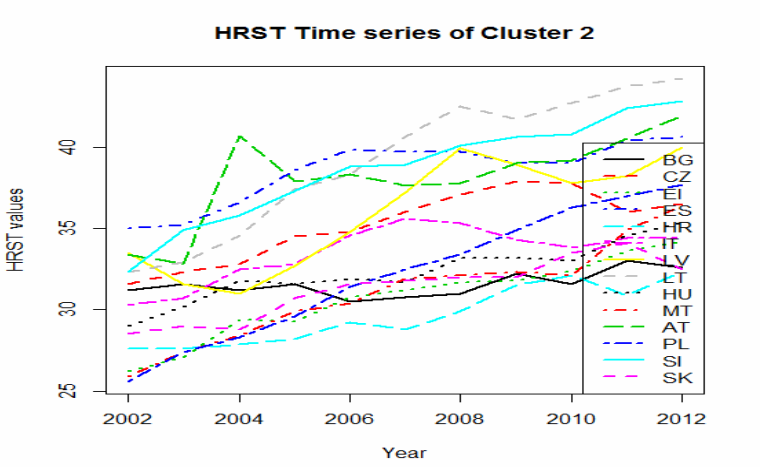


**Fig. 6.** HRST Time series of Cluster 2 *Source:* made by the author in R

Table 3 presents the means and standard deviations of HRST indicator values for each country from Cluster 2.

**Table 3.** Means and standard deviations of HRST indicator values for the countries from Cluster 2

| Country[2] | BG | CZ | EL | ES | HR | IT | LV | LT | HU | MT | AT | PL | SI | SK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 31.6 | 35.2 | 30.7 | 38.5 | 29.6 | 33.5 | 35.1 | 39.2 | 32.3 | 31.1 | 38.1 | 32.2 | 38.6 | 31.3 |
| Std Dev | 0.8 | 2.2 | 2.5 | 1.1 | 1.8 | 1.7 | 3.3 | 4.3 | 1.8 | 3.1 | 2.8 | 4.1 | 3.3 | 1.9 |

Source: made by the author in R

For cluster 3 we notice the fact that it is very well marked (average silhouette=0.89) and it comprises only two countries: Romania and Portugal (see Figure 7).

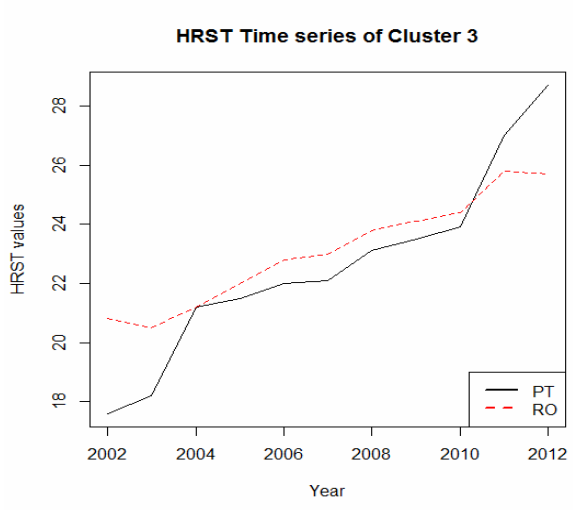


**Fig. 7.** HRST Time series of Cluster 2 *Source:* made by the author in R

[2] Europe ISO country code (ISO-3166-2) available at http://www.countrycallingcodes.com/iso-country-codes/europe-codes.php

Table 4 introduces the means and standard deviations of HRST indicator values for each country from Cluster 3.

**Table 4.** Means and standard deviations of HRST indicator values for the countries from Cluster 3

| Country[3] | BG | CZ | EL | ES | HR | IT | LV | LT | HU | MT | AT | PL | SI | SK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 31.6 | 35.2 | 30.7 | 38.5 | 29.6 | 33.5 | 35.1 | 39.2 | 32.3 | 31.1 | 38.1 | 32.2 | 38.6 | 31.3 |
| Std Dev | 0.8 | 2.2 | 2.5 | 1.1 | 1.8 | 1.7 | 3.3 | 4.3 | 1.8 | 3.1 | 2.8 | 4.1 | 3.3 | 1.9 |

Source: made by the author in R

## Conclusions

In this paper we proposed a classification of the EU countries relying on their similar evolution of HRST indicator during the period 2002-2012. Employing the clustering time series method we highlighted a structure composed of three clusters. The quality of the obtained structure, accounted for via Silhouette Coefficient value and by means of the silhouettes graphic representation, can be appreciated as reasonable, taking into account the results from the extant literature.

Finally, by analyzing the data from tables 2, 3 and 4, we can characterize the obtained clusters as follows:

- In cluster 1 some EU countries are grouped, countries whose evolution is distinguished by *high* means of HRST indicator values, between 42.1% and 49.8% (see table 2);
- The countries grouped in cluster 2 (half of the EU countries) are distinguished by *moderate* means of HRST indicator values, between 29.6% and 38.6% (see table 3);
- Cluster 3 groups other EU countries whose evolution is distinguished by *modest* means of HRST indicator values, in particular 22.6% for Portugal and 23.1% for Romania (see table 4).

## References

1. Hastie, T., Tibshirani, R., Friedman, J., T*he elements of statistical learning, - Data mining*, Inference and Prediction, Springer-Verlag, New York, LLC, 2003
2. http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/priorities/index_ro.htm
3. http://epp.eurostat.ec.europa.eu/portal/page/portal/science_technology_innovation/data/main_tables
4. http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_769.pdf
5. http://www.jstatsoft.org/v01/i04/paper

[3] Europe ISO country code (ISO-3166-2 ) available at http://www.countrycallingcodes.com/iso-country-codes/europe-codes.php