# Financial Development Indicators: A Comparative Study between Lebanon and Middle East Countries Based on Data Mining Techniques

Souha el Katat[1], Ali Kalakech[1], Mariam Kalakech[1], and Denis Hamad[2]
[1]Faculty of Economics and Business Administration, Lebanese University, Lebanon
[2]Laboratoire d'Informatique, Université du Littoral Côte d'Opale, France

**Abstract:** *Fighting poverty is one of the main objectives of sustainable development program. In a country like Lebanon, where poverty is a real threat and hidden under a good living looking, the situation should be explored in depth. This paper aims to evaluate the position of Lebanon compared to other Middle East countries in sustainable development. Furthermore, our goal is to reveal the power and weaknesses of resources management, based on income and non-income indicators retrieved from World data bank. For this purpose, we adopted a combination of data mining techniques as tools to study the relationship between these indicators. The K-means clustering technique is used to define the different levels of living. In order to extract the most relevant non-income indicators to our study, information gain as feature selection technique was applied. Finally, k-Nearest Neighbor (KNN) classification technique was used for the predicting model.*

**Keywords***:* *KNN, information Gain, K-means, world development indicators, sustainable development.*

*Received September 27, 2018; accepted January 22, 2019*

## 1. Introduction

Since 2000 till our present date, Lebanon went through many ups and downs, wars on several fronts and crisis in different kinds. Yet, a young Lebanese looks literate and having a dynamic social life, on the other hand he dreams of buying a house and travelling for working outside the country is first plan after graduation.

In numbers, according to United Nations Development Program (UNDP) and Central Administration of Statistics of Lebanon, poverty rate in December 2015 is estimated at 27%, and reaches 38% in some regions, unemployment rate lately exceeded 37%, whereas the classification of Lebanon is considered among the highest human development countries, since its Human Development Index (HDI) is equal to 0.763 [11].

Being a part of the Middle East region, we decided to study the situation of Lebanon in comparison with other Middle Eastern countries using data mining techniques, to evaluate its performance compared to his entourage. For this purpose we used The World Development Indicators (WDI). These indicators are several measures taken by world data bank to evaluate the performance and the progress of the economies and countries since 1956 till 2016 in Sustainable Development [9].

The objective of this study is, first, to compare Lebanon's situation among the other Middle Eastern countries and find to which class or living level it belongs, based on its financial indicators. Second, we aim to define the most relevant non-financial living indicators that affect this classification, and thereafter, to know how homogenous Lebanon is with its neighborhood.

Thus, the WDI were separated into income and non-income, and the wanted is to study the effect of non-income on income. To realize our goal, we decided to use data mining techniques as follows:

First, we used k-means clustering techniques, to find the classification of Lebanon and its situation compared to the other Middle Eastern countries. In this part the living level of each of the countries will be defined.

Second, the info gain feature selection will be used in order to select the most influential non-financial indicators on the financial classification found in the previous part.

Finally, we used k-Nearest Neighbor (KNN) classification based on the selected indicators, the Middle East non-income data with the cluster label are taken for a prediction model: and considered Lebanon as a testing model, removing its cluster label and try to predict it, trying to find if indicators affecting Middle East countries will affect Lebanon's financial classification.

This paper will be organized as follows: in section 2, we start by listing few related works with our topic, in section 3we will display the data representation. Then, we define the KNN algorithm and the information gain feature selection technique. In section4, we exploit the data of our case study. Finally in section 5, we finish with the conclusion and discussing the results and findings.

## 2. Related Works

Actually, the confusing classification of the country, and the fact that no analysis was made with these indicators in Lebanon using data mining techniques, only few statistical reports were made, were the reason behind this study [6].

Nevertheless, several studies merged economic indicators with data mining techniques for some statistical facts and data description, and new features extractions and predictions.

Popescu and Andreica [8] found that within 28 E.U. countries, labor productivity indicator determines whether the country is classified among high economic performance countries or lower economic performance. For this aim, he applied Hierarchical Cluster analysis as a first step, and then a method based on CHAID classification trees.

Divya and Agarwal [3] used other technique to classify 70 countries. The investigated dataset contains macro-economic indicators such as Economic Freedom Index (EFI) and HDI, It allowed the author to divide the countries into three groups; developed, developing and under development countries.

Another study realized by Cheng [2] applied k-means cluster algorithm in regional economy for a comparative analysis on 30 areas of China. His study aimed to consider dynamically the influence of natural resources on Chinese regional economy, as well as the influence of development on regional economy.

## 3. Data Representation

To represent our data, we consider the representative matrix of *n* rows and *d* columns $X=(x_i^r),i=1.....n; r=1....d$; where

- *d* is the number of studied world development indicators,
- *n* is the number of (countries x years),
- $x_i^r$ is the $i^{th}$ WDI of the $r^{th}$ Middle East country,

$$X=\begin{bmatrix} x_1^1 & \dots & x_1^r & \dots & x_1^d \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^r & \dots & x_i^d \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^r & \dots & x_n^d \end{bmatrix}$$

Each row

$$x_i=(x_i^1,\dots,x_i^r,\dots x_i^d)\in\mathbb{R}^d$$

Represents the indicators values for a specific country in a specific year, whereas each column

$$f^r=(x_1^r,\dots,x_i^r,\dots x_n^r)^T\in\mathbb{R}^n$$

Represents the values of one indicator for the different considered countries during the period of the study.

The classification that will be added for each country in every year is represented by the vector Y

$$Y=\begin{bmatrix} y_1 \\ \dots \\ y_i \\ \dots \\ y_n \end{bmatrix}$$

In a way for each country in every certain year, a $y_i$ value will be associated as a label, and this value is considered as the class of this country for this specific year.

Considering a high number of indicators in this matrix requires large storage capacity and computing time. That is why it is unavoidable to reduce this number of indicators. In our case, we chose to select the most relevant indicators that affect the classification of the countries, and that by using the information gain method for feature selection.

### 3.1. K-means Algorithm for Clustering

K-means algorithm is known for being one of the simplest clustering algorithms. The concept of this algorithm is to divide the given data into a predefined number of groups i.e., k groups or k clusters.

Each cluster has a center, far from the other centers. Each center defines the characteristics of the data belonging to this cluster, therefore, it helps distributing every other record in the nearest convenient cluster.

The function of partitioning for the objects ($x_1$, $x_2$, …, $x_n$), and *n* points that should be distributed in *k* clusters, it calculates the distance between the point and the centroid, and assigns the object to the smallest distance.

$$J=\sum_{j=1}^{K}\sum_{i=1}^{n}\left\|x_i^{(j)}-c_j\right\|^2 \quad (1)$$

Where $\left\|x_i^{(j)}-c_j\right\|^2$ is the distance between a point $x_i^{(j)}$ and the centroid $c_j$.

Once all the objects are distributed, it recalculates the position of centroids until the *k* clusters centers are not changing.

### 3.2. Information Gain for Indicator Selection

The information gain is a method commonly used for supervised feature selection. We use this method for indicator selection. It measures the reduction of entropy which reflects the amount of information gained by knowing the value of the indicator [1].

$$\text{Info Gain }(f^r)=H\,(\text{Class})-H\,(\text{Class}\mid f^r). \quad (2)$$

Where *H* refers to entropy is:

$$H(X)=-\sum_{i=1}^{n}P(f^r).log_2\,P(f^r) \quad (3)$$

For each indicator, one info gain value is calculated. The more it gets closer to 1 (maximum value), the more the indicator affects the information and considered with stronger power to classify the data, as well, the more it gets near to 0 (minimum value) the less the indicator holds information, and, then it can be disregarded [5].

# 4. Case Study

## 4.1. Data Source

The data we are using is obtained from the World data bank. This latter provides data created to evaluate the quality assurance in many domains concerning 189 countries, form 1962 to 2016.

The selected data for this study concerns the indicators of the 16 Middle Eastern countries, from 2000 till 2016.

The concerned countries are Bahrain, Cyprus, Egypt, Iran, Iraq, Palestine, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, and Yemen.

## 4.2. Preprocessing and Normalization

As a first step, the data should be cleaned by:

- Eliminating data in years 2002-2003-2004, where most indicators values were missing as well as for the years 2015 and 2016.
- Eliminating the indicators that was mostly empty.
- Filling the missing values by approximate values, especially when these values belonged to linear shapes, or average values when curve shape was not well defined [7].

Second step was to understand data and select the needed indicators. The chosen indicators are meant to be separated into two groups:

- *Income data*: a group of 57 indicators that reveal the living level of the country. This group contains economy, technology, social development, work, energy and finance indicators.
- *Non-income data*: the group that contains all the indicators concerning poverty, health, education and urban development domains. This group contains 56 indicators that reflect the living quality and resources of the country.

Indicators covering other domains that don't match our study (like environment, climate…) were removed.

The third step in data preprocessing is to unify the weight of all the indicators. In this step, data must be normalized, in order to have all the values between 0 and 1.

For this purpose we used linear normalization using the formula below:

$$z = \frac{x - min(x)}{[max(x) - min(x)]} \tag{4}$$

Both data groups are then represented in the two matrices *X1* and *X2* below:

*X1* is the in Income data matrix, defined as:

$$X1 = \begin{bmatrix} x_1^1 & \dots & x_1^r & \dots & x_1^{d1} \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^r & \dots & x_i^{d1} \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^r & \dots & x_n^{d1} \end{bmatrix}$$

Where: *d1*= 57, the number of income indicators
*n* = 16 countries x 12 years = 192 rows

*X2* is the Non-income data matrix defined as:

$$X2 = \begin{bmatrix} x_1^1 & \dots & x_1^r & \dots & x_1^{d2} \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^r & \dots & x_i^{d2} \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^r & \dots & x_n^{d2} \end{bmatrix}$$

Where: *d2*=56, the number of Non-income indicators
*n* = 16 countries *x* 12 years = 192 rows

## 4.3. Clustering

Clustering is a non-supervised technique. It is used to assemble similar records in one group named cluster. The data found in one cluster are homogeneous whereas heterogeneous with data found in other clusters. This grouping is done automatically by algorithms chosen based on data and clusters specification [4].

For this purpose, we chose the K-means algorithm to cluster the different countries on yearly basis. Since the UNDP usually classifies the world's countries into four categories of development, we used then the same number of clusters. The clustering of the countries was in four classes each representing the living level of the country [10].

By distributing the countries over clusters, based on their income indicators, the belonging cluster will be considered the label of every country. The label will be added to the non-income indicator, in order to choose best resources or non-income indicators affecting the financial status of the countries in a later step.

In this part of the study, we need to find the countries classification. We used the income data of the matrix *X1* to classify the countries.

The result of this step that came out was as the Table 1 and Table 2 below:

Table 1 shows the distribution of the Middle East countries on the four clusters. Each cluster shows a separate living level. We notice that the number of countries is not the same over the years.

To investigate in depth the behavior of each country over the considered year, we detail the distribution in Table 2.

Table 1. Number of countries in every cluster on yearly basis in the Middle East Countries based on their Income Indicators.

| | **2000** | **2001** | **2005** | **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cluster 0** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **Cluster 1** | 5 | 5 | 4 | 4 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 |
| **Cluster 2** | 3 | 3 | 5 | 6 | 8 | 7 | 7 | 4 | 5 | 3 | 3 | 5 |
| **Cluster 3** | 7 | 7 | 6 | 5 | 5 | 6 | 6 | 9 | 8 | 9 | 9 | 7 |

According to the Tables 1 and 2 we can clearly notice how the different countries were distributed over the four clusters, each representing the financial level of the belonging countries:

- *Cluster 0*: is the lowest living level. This cluster was all the time reserved to Yemen who lives in extreme poverty during all the years of this study. The indicators of Yemen were extremely low when it should be high, and conversely.
- *Cluster 1*: Low living level. The countries detected in this cluster are the countries living in bad conditions, some in war and some in poverty, such as Egypt or Iraq…
- *Cluster 2*: Medium living level. Countries in this level moved in and out of the level, depending on the circumstances of each country sometimes getting into a better living level, and sometimes not. In other words, upraising from developing country to developed county.
- *Cluster 3*: High living level, where we find almost the half of the Middle East countries, especially in the latest years of the study. A positive development sign is shown by the increasing number of countries in this cluster.

Table 2. The distribution of the Middle East Countries in clusters based on their Income Indicators.

| **Country** | **Yearly distribution over clusters** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **2000** | **2001** | **2005** | **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **2012** | **2013** | **2014** |
| **Yemen, Rep.** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **United Arab Emirates** | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| **Turkey** | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **Syrian Arab Republic** | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| **Saudi Arabia** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **Qatar** | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| **Oman** | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| **Lebanon** | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| **Kuwait** | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| **Jordan** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 3 | 2 |
| **Palestine** | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 2 |
| **Iraq** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **Iran, Islamic Rep.** | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **Egypt, Arab Rep.** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **Cyprus** | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| **Bahrain** | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

These clusters are then considered as the class label*y* for each one of the countries every year. A vector to be added to the non-income data represented by the *X2* matrix:

$$X2 \ + \ Y \ = \begin{bmatrix} x_1^1 & \dots & x_1^r & \dots & x_1^{d2} \\ \dots & \dots & \dots & \dots & \dots \\ x_i^1 & \dots & x_i^r & \dots & x_i^{d2} \\ \dots & \dots & \dots & \dots & \dots \\ x_n^1 & \dots & x_n^r & \dots & x_n^{d2} \end{bmatrix} + \begin{bmatrix} y_1 \\ \dots \\ y_i \\ \dots \\ y_n \end{bmatrix} \qquad (5)$$

Where *n* is the number of (countries *x* years) = 192 rows, *d2* = 57 Non-income indicators

And *Y* is the cluster label vector obtained by the previous clustering technique.

## 4.4. Indicators Selection

After labeling our set of data, and in order to build the most accurate classifier, we need to select the most representative income indicators from the input space. In fact, previous works have shown that selecting the most discriminative inputs may lead to better results.

To choose the most relevant non- income indicators affecting and having strong effect on the financial level of the Middle Eastern countries, the Information gain indicator selection was applied.

In order to measure the efficiency of those selected indicators, we use the accuracy rates of the KNNclassifier that operates on the space defined by the indicators selected by the Info gain method.

At each step, one Indicator is added at the time and the accuracy is evaluated until considering the whole income indicators [12].

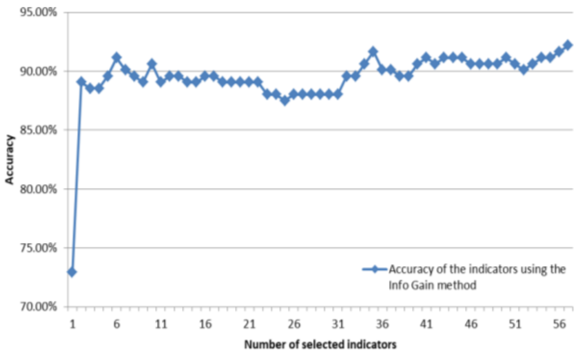

Figure 1. Accuracy of the Info Gain method.

Figure 1 shows the plots of accuracy vs. the desired number of non-income selected indicators by the Info gain method. From this figure, we can see that the accuracy of the first Indicator starts with 72.92%. This accuracy increases slowly to reach a maximal value of 91.15% when the first six Indicators are selected.

When d>6, this accuracy rate became stable or even decreases when we select new indicators. This clearly confirms the interest of the selection procedure before performing the classification.

Since the highest accuracy rate is obtained by selecting the first 6 indicators, we present below a list of those indicators:

- Labor force participation rate for ages 15-24 female (%).
- Aquaculture production.
- Unemployment female (% of female labor force).
- Total labor force participation rate (% of total population ages 15+).
- GDP per person employed.
- Male labor force participation rate (% of male population ages 15-64).

## 4.5. Classification of Lebanon

In order to study the situation of Lebanon, the six first selected indicators are used as inputs, and the labels obtained in clustering part are considered as outputs to build the prediction model.

The prediction of Lebanon classification obtained by the prediction model on yearly basis is shown in the Table 3, compared to the labels obtained earlier by the clustering.

Table 3. Comparison between the classifications of Lebanon based on non-income indicators and the income indicators based clustering.

| Year | Classification based on non-income selected indicators prediction model | Clustering based on income indicators |
|---|---|---|
| 2000 | 1 | 3 |
| 2001 | 1 | 3 |
| 2005 | 1 | 2 |
| 2006 | 1 | 2 |
| 2007 | 1 | 2 |
| 2008 | 1 | 2 |
| 2009 | 1 | 2 |
| 2010 | 1 | 3 |
| 2011 | 1 | 3 |
| 2012 | 1 | 3 |
| 2013 | 1 | 3 |
| 2014 | 1 | 3 |

Table 3 clearly shows the wide difference between the classifications of Lebanon based on non-income indicators and the income indicators.

In the clustering based on income indicators, Lebanon was considered to be among the high ranked countries (cluster 2 and cluster 3), which means Lebanon has high income. When it came to the selected non-income indicators that affect the most on the ranking of the other Middle East countries, Lebanon should be classified in the low living level (cluster 1).This difference might be due to many options and conditions, stated in the conclusion.

The model could be summed up as follows in Figure 2.

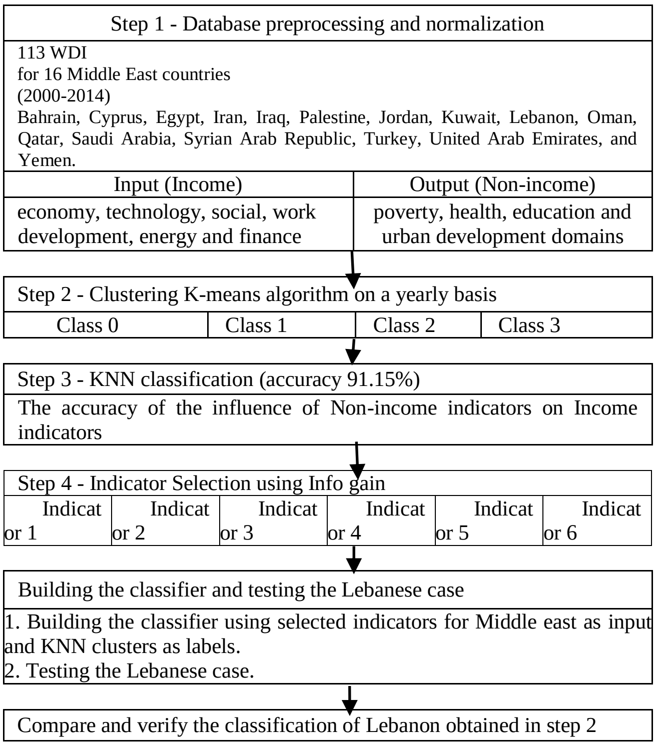
Step 1 - Database preprocessing and normalization
113 WDI
for 16 Middle East countries
(2000-2014)
Bahrain, Cyprus, Egypt, Iran, Iraq, Palestine, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, and Yemen.
Input (Income)
Output (Non-income)
economy, technology, social, work development, energy and finance
poverty, health, education and urban development domains
Step 2 - Clustering K-means algorithm on a yearly basis
Class 0
Class 1
Class 2
Class 3
Step 3 - KNN classification (accuracy 91.15%)
The accuracy of the influence of Non-income indicators on Income indicators
Step 4 - Indicator Selection using Info gain
Indicator 1
Indicator 2
Indicator 3
Indicator 4
Indicator 5
Indicator 6
Building the classifier and testing the Lebanese case
1. Building the classifier using selected indicators for Middle east as input and KNN clusters as labels.
2. Testing the Lebanese case.
Compare and verify the classification of Lebanon obtained in step 2

Figure 2. Flowchart.

## 5. Conclusions

In this paper, our objective was to study the following points:

- The classification of Lebanon in comparison with other Middle Eastern countries based on its income indicators.
- The most relevant indicators affecting the classification.
- The homogeneity of Lebanon with his neighborhood.

The application of the combination of data mining techniques stressed the difference between the livings of the Middle East countries, and showed the strength of the participation and activity of the population in development. As well, it showed that Lebanon is displaced, and that it has other strength and weaknesses to be restudied.

First, we found that Lebanon is considered among the middle and in some years among high living level countries. But when we found the most relevant indicators, and we reclassified Lebanon based on these indicators, we obtained that Lebanon should have a bad living level.

Second, concerning the most relevant indicators, the results show the importance of the activity and the work of different parts of the society (males and females), and the aquaculture production, which might be interpreted by the neighborhood of all Middle East countries with oceans and seas. In fact all these indicators show one important issue, which is the activity of the population and the participation of all community layers in the development of their countries.

Concerning the third point for the evaluation of Lebanese situation, the results lead us to one or more of the following conclusions:

- Chosen indicators in the cleaning stage may not be the best set of representative indicators, or where there was some information regarding Lebanon, there was missing information for other countries. This leaded to disregard some significant indicators.
- Lebanon is not similar to other Middle Eastern countries, and the indicators that are relevant to other countries, does not really count in Lebanon.
- The Lebanese people may count on other resources than the chosen income indicators for a good living level. For example, debts are one of the main resources of living in Lebanon, either public debt which reached $76.17 billion in January 2017 or individual loans. One more resource that most Lebanese count on, is the money sent by relatives living out of the country.

In fact, much more analysis and techniques of data mining could be applied to explore the facts in depth, hence identify the strength and weaknesses in the Lebanese system and how to enhance the economy of the country.

## 6. Acknowledgment

This work is totally funded by the National Council for Scientific Research of Lebanon (CNRS-L), and we are very thankful.

## References

[1] Augusto V., http://www.emse.fr/~augusto/enseignement/icm/gis1/UP3-2-Fouille_de_donnees-handout.pdf, Last Visited 2017.

[2] Cheng Z., "Regional Economic Indicators Analysis Based on Data Mining," *in Proceedings of 5thInternational Conference on Intelligent Systems Design and Engineering Applications*, Hunan, pp. 762-730, 2014.

[3] Divya T. and Agarwal S., "Classification of Countries on Macro-economic Variables Using Fuzzy Support Vector Machine," *International Journal of Computer Applications*, vol. 27, no. 6, pp. 41-44, 2011.

[4] Fessant F., http://www.vincentlemaire-labs.fr/cours/2.2ApprentissageNonSupervise.pdf, Last Visited 2018.

[5] Han J., Kamber M., and Pei J., *Data Mining Concepts and Techniques*, Elsvier, 2012.

[6] Kalakech A., Hamad D., and Kalakesh M., "Selection of World Development Indicators for Countries Classification," *in Proceedings of International Conference on Digital Economy*, Carthage, pp. 24-28, 2016.

[7] Li J., Koronios A., and Natarajan K., "Data Mining Techniques for Data Cleaning," *in Proceedings of the 4th World Congress on Engineering Asset Management*, Athens, pp. 796-804, 2009.

[8] Popescu M. and Andreica M., "A Method to Improve Economic Performance Evaluation Using Classification Tree Model," *European Journal of Business and Social Sciences*, vol. 3, no. 4, pp. 249-256, 2014.

[9] UN, http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&referer=/english/&Lang=F, Last Visited 2018.

[10] UNDP, http://hdr.undp.org/sites/default/files/hdr2015_technical_notes.pdf, Last Visited 2018.

[11] UNDP, http://hdr.undp.org/sites/default/files/2016_human_development_report.pdf, Last Visited 2018.

[12] Wu x. and Kumar V., *the Top Ten Algorithms in Data Mining Book*, a Chapman and Hall/CRC, 2009.

**Denis Hamad** is professor at the University of Littoral Côte d'Opale. He obtained a HDR (Habilitation à Diriger la Recherche) degree in neural networks for unsupervised pattern classification and a PhD degree in detection and validation of measurements in complex systems from the University of Lille-France in 1997 resp. in 1986. Between 1998 and 2002, he was Professor at the University of Picardie Jules Vernes, Amiens-France. His main research interests are machine learning, image and signal processing.

**Ali Kalakech** is a professor at the Information Systems Department in the Lebanese University, Faculty of Economics and Business Administration. He got his Master Degree in Computer Systems from the National Institute of Applied Sciences, Toulouse, France in 2001. He received his Ph.D. degree from the National Polytechnic Institute, Toulouse, France in 2005. His Research interests include Data Mining, dependability of computer systems, networking and performance evaluation.

**Mariam Kalakech** is anassociate professor at the Information Systems Department in the Lebanese University, Faculty of Economics and Business Administration. She got her Master Degree in Information Systems from the National Institute of Applied Sciences, Lyon, France in 2007. She received her Ph.D. degree in Automatics and Industrial Computingfrom the University of Lille-France in 2011. HerResearch interests include machine learning, image and signal processing.

**Souha el Katat** obtained her bachelor in Computer engineering from CNAM University – Beirut, and then completed the Master degree in Business Computer from the Lebanese University – Faculty of Economics and Business Administration. Her research interests include Data Mining and machine Learning.