Research Article

Solar Energy Validation for Strategic Investment Planning via Comparative Data Mining Methods: An Expanded Example within the Cities of Turkey

Oya H. Yuregir[1] and Cagri Sagiroglu[2]

[1]*Faculty of Engineering and Architecture, Industrial Engineering, Cukurova University, 01330 Adana, Turkey*
[2]*Faculty of Arts, IRIO Department, Groningen University, 9712 EK Groningen, Netherlands*

Correspondence should be addressed to Cagri Sagiroglu; cagri.sagiroglu@rug.nl

Energy supply together with the data management is one of the key challenges of our century. Specifically, to decrease the climate change effects as energy requirement increases day by day poses a serious dilemma. It can be adequately reconciled with innovative data management in (renewable) energy technologies. The new environmental-friendly planning methods and investments that are discussed by researchers, governments, NGOs, and companies will give the basic and most important variables in shaping the future. We use modern data mining methods (SOM and K-Means) and official governmental statistics for clustering cities according to their consumption similarities, the level of welfare, and growth rate and compare them with their potential of renewable resources with the help of Rapid Miner 5.1 and MATLAB software. The data mining was chosen to make the possible secret relations visible within the variables that can be unpredictable at first sight. Here, we aim to see the success level of the chosen algorithms in validation process simultaneously with the utilized software. Additionally, we aim to improve innovative approach for decision-makers and stakeholders about which renewable resource is the most suitable for an exact region by taking care of different variables at the same time.

1. Introduction

Today, concerns about the environment, energy security, and economic prosperity are increasingly developing. The International Energy Agency (2013-2014) estimated that demand for primary energy will increase globally by 55 percent between 2005 and 2030. Additionally, in developing countries where economies and population are expected to grow fast and primary energy demand is projected to grow by 74 percent during the same period, fossil fuels are expected to remain the dominant source of primary energy, accounting for 85 percent of the overall increase in global demand [1]. But the question is how long this unsustainable de facto situation can continue.

It is observable that since the 1990s, the attempt to convert production systems to clean production systems has not been the aim of the companies alone, but this was started to be implemented by joint operations between the industry, government bodies, and the public [2]. In addition to be a key criterion of the sustainable development stipulation, for business and states/interstates, the integration of social and ecological challenges by managing the data in an innovative manner into decision-making process from the beginning is of growing importance for minimizing future environmentally related risk and sustainable associated economic costs [3–5].

In this perspective, this research focuses on verification and experimental comparison of the two possible data mining options (and data mining itself) for new solar energy investment decisions. In Section 2, we gave the basic background in multivariate cluster analysis, including Self-Organizing Map and K-Means with software tools for comparison. Additionally, we explained why these methods have been chosen. Our data owing to the cities have been gathered

TABLE 1: Data analysis task and techniques [6].

Data analysis techniques	Data summarization	Segmentation	Classification	Prediction	Dependency analysis
Descriptive and visualization	•	•			•
Correlation analysis					•
Cluster analysis		•			
Discriminant analysis			•		
Regression analysis				•	•
Neural networks		⊙	⊙	⊙	
Case-based reasoning					•
Decision trees			•	•	
Association rules					•

from governmental/special official statistics foundations via mail and/or face to face conversations. In Section 3, we commanded, compared, and discussed the graphics about solar radiation and consumption values which are created by Rapid Miner 5.1 and MATLAB. These programs have been used for comparison of the results, respectively. In Section 4, we summarized the results and we proposed our suggestions about algorithms in pattern recognition for validating the investment decision. As a result, in addition to present new approach for assessment of the (future) solar power investments, we will suggest the better algorithm and software for decision-making and optimization process for stakeholders.

2. Material and Methods

2.1. Data Mining and Clustering. The objective of data mining (DM) is to identify valid novel, potentially useful, and understandable correlations and patterns in existing data and finding useful patterns in data is known by different names in different communities [8, 9]. Chung and Gray and Liao and Triantaphyllou gathered the recent updates of data mining algorithms in different disciplines and diversified range of applications [10, 11]. Also from the ecology point of view, various successful studies were collected as examples of different applications [12]. While many data mining tasks follow a traditional, hypothesis-driven data analysis approach, it is commonplace to employ an opportunistic, data-driven approach that encourages the pattern detection algorithms to find useful trends, patterns, and relationships [6]. In this manner, as shown in Table 1, we decided the neural networks as the best solution for our study since it is important from the relationship point of view. Because of a modern analyzing studies complication and hardness to confirm (or validate) the results, it is better to use more than one method with the same variable attributes.

The similar trilemma of energy/data/optimization has been used to analyze clustering of 316 different European regions according to their cultural background [13]. From the other side, Seidler and Adderley used data mining approach to develop a proactive approach against the criminal networks [14]. Additionally, Santillana et al. [15] present a methodology mainly based on data mining and machine learning techniques not only for observation but also for estimation and evaluation of the influenza disease effects.

Clustering, as the main methodology of this study, is *simply* division of data into groups of similar objects [8]. There are so many methods and algorithms to do that in the literature. But, two different clustering algorithms are chosen to investigate the data of this study and compare their results: Self-Organizing Maps and K-Means algorithms. Sobkowicz et al. (2012) [16], similar to us, used successfully these algorithms for modeling, simulating, and forecasting of the ideas in the web and social media. The general reasons for selecting these two algorithms are their popularity, flexibility, applicability, and handling high dimensionality [17].

Additionally, Figure 1 shows total electricity cost including construction, production, and decommissioning for renewal energy. Published in September 2014, the chart shows that electricity from hydro power is already more cost effective than electricity from other fuels. Even the price for electricity from wind power is becoming competitive in some cases with that of natural gas and almost with that of coal-generated energy. The price of power from solar panels is still high, but the cost differences decrease day-by-day mainly due to more efficient panels, mass production, and unstable fossil fuel costs. Many studies indicate that the transition cost of the energy system will be far less than the long term cost for keeping on using the fossil fuels. Since the most expensive energy is the still solar, the investment decision must be analyzed very carefully from various perspectives. In this study, we chose the solar energy values because it is the widest one and simple to build nearly all around the world.
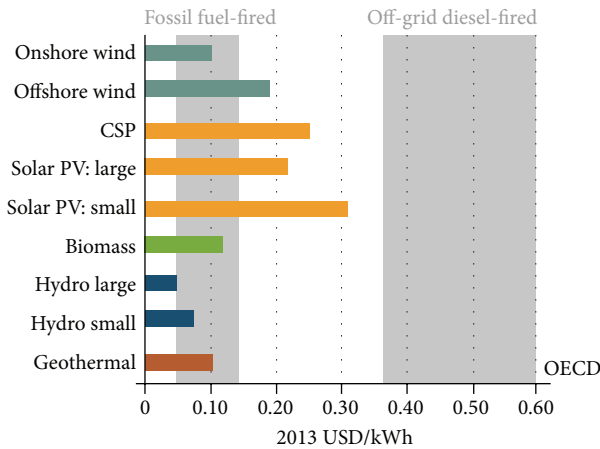
FIGURE 1: Total electricity cost (source: International Renewable Energy Agency, REthinking Energy: Towards a new power system, 2014) [7].

We used clustering for exposing the secret relationship among consumption similarities of cities for making cost effective investments. Therefore, it is expected to obtain new neighborly relations that will be the base of the new renewable source investments according to the new clustered structure.

2.2. Self-Organizing Maps.
Self-Organizing Maps (SOM) were presented in the beginning of the 1980s by Kohonen, Finland academician. The main idea of SOM is to map the data patterns onto an n-dimensional grid of units or neurons [18] and it is usually used for mapping high-dimensional data into one-, two-, or three-dimensional feature maps (simple geometric relationships) to increase intelligibility [19, 20]. Additionally, the SOM is an adaptive display method that is particularly suitable for the representation of structured/normalized statistical data. The mapping represents a data set in an ordered form, whereby mutual similarities of data samples will be visualized as geometric relations of the images of the samples on the map [21]. SOM procedures are used in a range of applications, but they are having a major impact in the fields of data exploration and data mining [22].

Traditional multivariate statistical approaches are often confused by data sets with variable relationships that are nonlinear, by data distributions that are abnormal (typically with multiple populations), and by the data sets themselves that may be disparate, sparsely filled (contain "nulls") with both continuous and discontinuous numeric data and text [23]. The SOM, ordered vector-quantization approach can overcome many of these problematic issues [24] and due to its layer-based structure effects of independent variables that can be visible separately. Lately, Miche et al. [25] adapted also SOM clustering a priori knowledge successfully in 5 different examples.

2.3. K-Means Algorithm.
The K-Means algorithm, one of the mostly used and well-known clustering algorithms, is classified as a partition or nonhierarchical clustering method. K-Means is typically used with the Euclidean metric for

computing the distance between points and cluster centers [26]. It quickly converges to a local minimum of its cost function [27]. In K-Means, because the centers are being connected to each other, the data in one- or two-dimension reflection in space can be achieved [28]. Additionally, K-Means was successfully used as a solution for accuracy challenge in large-scale image classification as a part of hybrid model that shows the adaptable structure of the algorithm into different fields like another example [29] cited by Elssied et al. [30] from soft computing perspective.

Briefly, the primary objective of applying K-Means is to provide groups whose members are close (have high similarity degree) and well separated. In the clustering process, there are no predefined classes and no examples that would show what kind of desirable relations should be valid among the data, so it is natural to be asked about the validity and quality of the results obtained [31]. At every step, the center of each cluster is recalculated by using the average vector of the objects (which are assigned to the same cluster) and every object can just be situated in one cluster [32].

2.4. System Architecture.
As mentioned in many studies, focusing on just consumption values to analyze energy challenge is not enough by itself. Other necessary influencing factors should be cared for as well. In this manner, first we grouped the cities in accordance with their electricity consumption. Under consumption, we have seven different variables. At least, we compared 81 cities with seven different subtitles of total consumption (i.e., government, industry, agriculture, city lighting, housing, trade, and others) and we also took care of data between 2004 and 2014 (totally 5670 values). An example data set with the explanation is presented in Appendix A. Our iteration coefficient to reach the truer results is 1000 times. We used SOM and K-Means (as a representative of the nonhierarchical methods) algorithms. Because we want to see the sensitiveness of the algorithms primarily with the limited data, we set the k value of SOM and K-Means equal and compared their performance when $k = 5$. (We picked 5 according to first observable outputs of randomly chosen data sets which were handled via MATLAB and Rapid Miner 5.1.)

Later, we normalized and clustered the data in SOM and then K-Means with (first) both of the software programs: MATLAB and Rapid Miner. There are more than one data mining tool as R, WEKA, Orange, Rapid Miner, Tanagra, KNIME, and so forth in addition to MATLAB which is well known tool especially in engineering and mathematical disciplines. We chose Rapid Miner as an alternative to MATLAB mainly because it is user friendly and with modern structure in ingredients for free. Rapid Miner also has the most important connectors: databases, Excel, txt, html, CSV, and so forth. In Rapid Miner, it is easy to use wizards for setting up your data sources and a graphical environment for processing data flows. Rapid Miner can use every algorithm in the other programs and has very innovative pathways to offer possible solutions and mistakes. Although WEKA looks the most successful data mining tool according to the study of Borges et al. (2013) [33], they did not take Rapid Miner
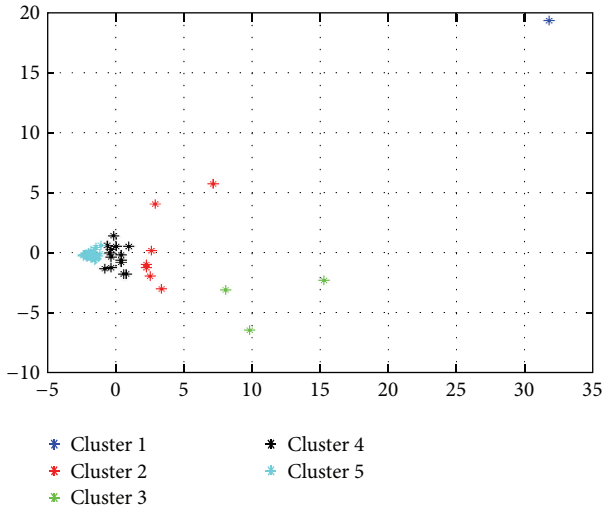
FIGURE 2: Clustering results.

FIGURE 3: The cities SOM neighbor weight distances (Rapid Miner 5.1).

as a candidate in their research. Generally Rapid Miner, R, Weka, and KNIME have most of the desired characteristics for a fully functional platform [34], but lastly Piatetsky [35] presented that in 2013 Rapid Miner was the mostly used open source tool in real projects.

First of all, we manipulate our (prepared) data with K-Means and SOM in MATLAB. However, we reduce the variables from seven to three to ensure especially duration and stability in MATLAB. Reduced data can be visualized in an easier way in two dimensions and it does not have meaningful effect on the results in case of using one of these methods: data aggregation, dimension reduction, data compression, and discretization [36]. During reducing variables, we calculated the most effective inputs (with standard variations) via principal component analysis method.

In this manner, Figure 2 shows the first clustering results on the graph briefly. The algorithm that has been used in MATLAB and the detailed results are also presented in Appendix B.

Next, *the same data* were handled via Rapid Miner 5.1. in addition to MATLAB for comparison. The table in Appendix C shows the clustering results separately according to consumption values by K-Means and SOM. From one perspective, it shows the relevance for investment requirements and the clusters.

After observing the clusters of similarities according to consumption (independent of geographical neighborhood), we decided to find similarities of their solar power potential. The detailed and sensitive data has been gathered from national and local meteorology authorities in addition to archives of Ministry of Energy. The summary of data and algorithm for MATLAB that has been utilized with the results are presented in Appendix D. Furthermore, the Rapid Miner 5.1. results for SOM and K-Means are in Appendix E.

These two data sets (solar radiation and consumption) are suitable to compare for observing which algorithm is more accurate and sensitive. In this manner, Figure 3 illustrates clustering schema of the cities according to SOM. Based on
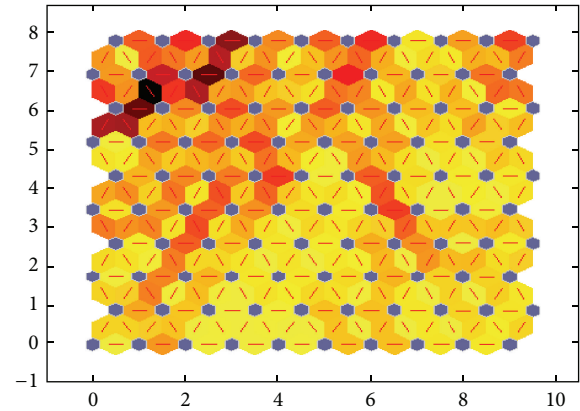
the profiles of the clusters identified by the SOM, we could say that SOM grouped the clusters successfully but not *too sensitively*. The yellow part spreads out remarkable area. Dark yellow and orange parts are represented with just a small area. But still, five different groups can also be observed from the spreading of the colors which has already been presented with details in Appendix F.

Till now, we have analyzed which cities are similar to each other from the point of electricity consumption and potential solar power views and also which algorithm gives better results. But for better decision support system (to serve the other aim of the study) we should know which city/cluster will need more power in the future simultaneously. Accordingly, we must know at least the increase of population and the ratio of industrialization. For example, if a city has got its own rich solar power potential but always loses its population or goes back in the welfare indexes because of any reason, it "strongly" means that for this city extra investment will not be necessary.

Because K-Means can provide more sensitive results versus SOM, from now on it has been decided to go on with K-Means via Rapid Miner 5.1. In this context, Figure 4 shows the clustering results according to population forecasting. The cities are clustered according to their similarity in population growth ratio and the results are visualized by color codes. The data have been supplied from Turkish Statistical Institute and can be found in detail in Appendix F. As seen, in every cluster, it is possible to find a city from any geographical region. So, it is possible to find "sister cities/regions" by focusing on the reasons in further studies. Additionally, there have been always some exceptions like megacities (Istanbul, Ankara, and Izmir in our example) where their values are far from the rest of the data and they can create separate research areas to themselves.

Similarly, Figure 5 shows the clustering of cities with color codes (and the numbers on the graph) in different way according to their welfare indexes values. Their index values have been provided by Ministry of Development and could be found in Appendix G. The colors represent the cities and the dimensions symbolize the index value. Smaller bubbles show
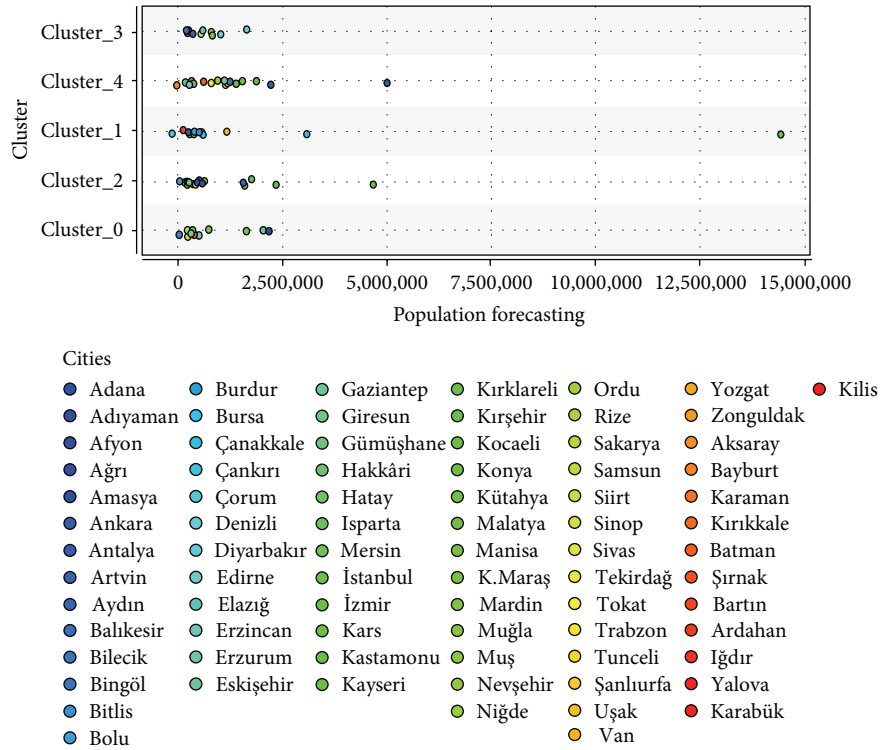
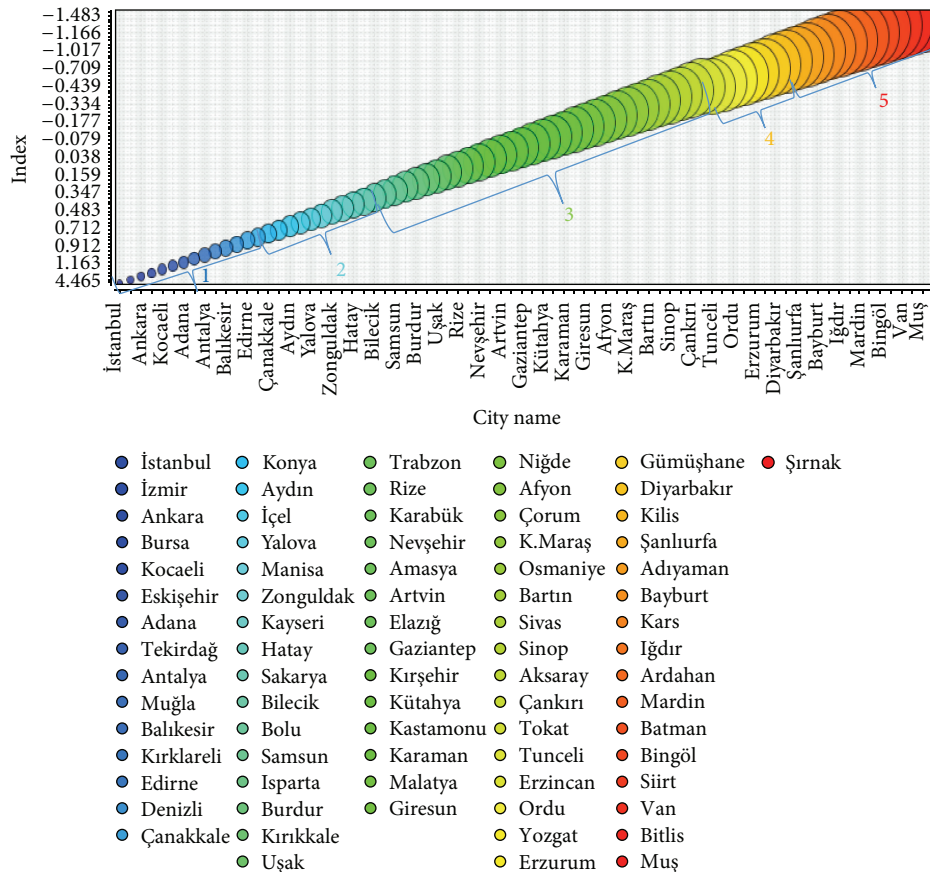FIGURE 4: Clustering results according to population forecasting.

FIGURE 5: Clustering cities according to their welfare indexes values.

more developed and bigger bubbles represent less developed cities. It is easy to notice that there is really big difference between the two extremes.

Therefore, it is possible to compare the results of cities' future energy demand indicators with their consumption-similarity neighborhood and their solar power neighborhood clusters. Decision-makers can see the basic decision support components in the integrity of a cluster.

3. Results and Discussion

As mentioned by World Bank and World Energy Council in 2015 Energy Trilemma Index [37, 38], energy-efficiency improvements have very positive developments in the field but they are still slow. Today's geopolitical as well as geoeconomic reality is dominated by urbanized areas, mainly organized as cities. The sustainability of cities, in particular megalopolises of over 10 million inhabitants, depends on regional and global networks for their basic needs in food, water, and energy. Security of supply and quality of life strongly depend on the functioning of these networks. But the present technology for producing electricity from renewable resources is still more expensive than the conventional ones despite technological development and various governmental supports. So, it is important to take care of cities/regions renewable richness not only by itself but also from different ways which at least should include the population forecasting, industry, agriculture and transformation rate and possibilities, climate, and the consumption routines. For this aim, both K-Means and SOM algorithms are suitable because of their convenience for huge and small data set, large and small number of clusters, and ideal and random data set usage and also feasible for different software alternatives. For nonoverlapping situations all the methods had good performance.

In our study, both algorithms were run with same data for similar number of iterations. Our result shows that K-Means is more sensitive and easier to implement. Additionally, when the k becomes greater, the performance of SOM decreases and when the data are not big enough, K-Means creates more successful outputs. In this context our findings are similar with Mingoti and Lima (2006) [23] and Abbas (2008) [39]. Likewise, Shahapurkar and Sundareshan [40] presented the results of application of the K-Means technique to data set which has 798 observations in 6 dimensions. According to their outcomes, K-Means clustering algorithm can provide various advantages compared to the SOM in terms of point density, accuracy, topology preservation, and computational requirements. Also, de Castro Leão et al. [41] showed in their study that at last K-Means can reach results as successful as SOM but in faster and simple manner which is partly similar with Bação et al. (2005) [42]. So, the paper shows that it is also possible to reach meaningful solutions without challenging huge amount of data via K-Means in a simple way. Additionally, Rapid Miner reached more sensitive results against MATLAB about clustering and forecasting in this paper with the same data.

Besides, we showed that data mining methods can be used as a successful tool also in renewable energy investment planning as many other areas. This study also illustrated in an innovative way that data mining methods successfully exposed the hidden relationship between the requirements and assets simultaneously by covering welfare ratio and cities energy requirements forecasting accordance to their growth of population and development. For instance, if we focus on the clusters in Figures 4 and 5, we realized that the cities that are at the same cluster according to their population potential *often* show the same welfare indexes line. On the other side, for a specific city/region, one can find its solar energy power, future need, and its welfare situation at one glance as a whole. If we focus on one city where we need more electricity in the future and have rich solar potential, it does not mean that we should invest sun-based production facility for sure. (Generally, the renewable resources are more suitable for home-based usage than the industrial ones because of their investment costs and own restrictions.) Briefly, the intersection(s) of these can give glue for the (more) accurate and efficient investment decisions.

4. Conclusion

In this paper, we have compared two clustering algorithms, respectively, via two different software programs to see which one is more successful under certain circumstances with sufficient (limited) data for the usage in new renewable energy investment decisions. In order to compare both algorithms, we have utilized four different types of data: electricity consumption, solar power potential, population forecasting, and welfare indexes of the cities. Although both software programs and algorithms are adequate enough for the analysis, from the time, easiness, and workforce perspectives, K-Means method and Rapid Miner tool have demonstrated their competence. This can help especially the social sciences researchers/decision-makers to check their theories in an easier way.

Additionally, we presented a validation system for administrators both in government and in private sector of energy management to keep track of the suitable areas of energy investment according to different factors such as region, population, and welfare. Particularly visualizing and analyzing the dynamics between the humanities and energy are innovative. This will offer new ways to improve insight in creating optimal strategies for sustainable smart energy systems. Thus, administrators can decide which city/cluster needs more energy and more importantly new energy investment. In this way, for future studies other factors affecting the energy consumption could be included for more specific results. This can be a start point of studies about nonconventional and *not state centric* but urban centric programming methods.

Appendix

A. Example Data Set (for 2004)

See Table 2.

TABLE 2: Example data for electricity consumption in sectors for 2004.

City	House	Commercial	Governmental	Industry	Agriculture	Street lighting	Others
Adana	27,7	12,3	3,2	44,2	1,9	3,4	7,2
Adiyaman	21,6	6,4	3,0	52,3	6,6	1,0	9,1
Afyonkarahisar	23,9	11,8	3,2	35,3	12,0	3,4	10,4
Ağri	50,2	8,8	12,0	5,7	0,0	7,0	16,2
Amasya	35,6	9,2	3,6	38,3	4,8	7,0	1,6
Ankara	33,6	22,3	15,1	24,7	1,7	2,2	0,4
Antalya	28,4	38,6	9,5	10,4	2,8	3,2	7,2
Artvin	29,9	15,1	9,2	42,0	0,0	3,6	0,2
Aydin	36,9	19,4	4,2	28,3	4,9	1,1	5,3
Balikesir	30,4	14,0	4,2	34,1	1,6	2,6	13,1
Bilecik	8,8	4,9	1,6	81,4	0,3	1,2	1,9
Bingöl	51,1	10,9	16,9	2,1	0,1	9,2	9,7
Bitlis	40,6	10,9	8,1	5,1	3,9	14,3	17,0
Bolu	16,6	17,0	3,6	51,5	0,3	3,1	7,8
Burdur	12,2	6,4	2,0	73,2	3,3	2,0	1,0
Bursa	17,3	10,5	1,8	65,5	0,6	1,0	3,3
Çanakkale	10,8	5,3	1,8	76,1	1,0	1,5	3,4
Çankiri	32,0	12,0	9,3	38,0	0,8	6,5	1,3
Çorum	30,7	11,6	3,2	41,7	2,2	4,3	6,3
Denizli	18,9	10,5	2,5	59,0	2,2	1,3	5,5
Diyarbakir	36,2	12,8	13,7	14,3	6,6	2,5	13,9
Edirne	26,1	12,5	4,2	34,0	10,4	2,8	10,0
Elaziğ	25,9	7,4	28,1	23,4	7,5	2,4	5,3
Erzincan	40,5	14,7	10,2	8,0	5,5	5,7	15,2
Erzurum	38,6	13,7	18,4	19,2	0,5	6,7	2,9
Eskişehir	21,9	10,0	6,3	47,0	5,0	2,8	6,9
Gaziantep	17,5	8,8	1,4	61,3	2,3	2,9	5,8
Giresun	51,3	13,3	5,8	14,9	0,0	9,5	5,1
Gümüşhane	51,0	14,6	9,1	5,2	1,5	10,3	8,3
Hakkari	33,5	7,3	22,9	0,4	0,0	30,8	5,1
Hatay	13,7	5,3	2,1	70,7	4,1	1,4	2,7
Isparta	19,4	3,8	7,3	61,0	4,8	2,0	1,7
Mersin	30,8	14,0	6,4	36,4	2,9	3,9	5,5
Istanbul	34,3	28,4	3,2	29,0	0,0	1,5	3,5
Izmir	20,1	9,9	2,7	58,9	2,1	2,4	3,9
Kars	27,1	9,9	9,3	22,7	0,1	10,5	20,4
Kastamonu	32,0	12,9	6,9	43,0	0,6	3,6	1,0
Kayseri	22,4	10,5	1,3	54,1	3,2	2,8	5,7
Kirklareli	12,1	5,0	1,5	74,1	0,1	0,9	6,2
Kirşehir	29,2	10,0	10,1	39,4	4,2	5,9	1,1
Kocaeli	8,6	6,6	1,8	81,7	0,0	0,7	0,6
Konya	17,4	8,8	4,5	47,8	16,9	2,5	2,1
Kütahya	24,0	10,9	3,6	52,9	0,3	3,5	4,7

TABLE 2: Continued.

City	House	Commercial	Governmental	Industry	Agriculture	Street lighting	Others
Malatya	27,4	12,8	7,6	36,6	6,9	5,9	2,8
Manisa	23,7	8,7	3,1	45,6	9,2	2,7	7,0
K. maraş	13,5	5,2	1,5	69,8	2,2	1,8	6,0
Mardin	14,3	5,5	7,6	27,8	23,0	5,4	16,4
Muğla	30,7	32,7	5,2	8,9	1,4	2,1	19,0
Muş	38,3	11,0	24,2	5,5	0,2	14,8	5,9
Nevşehir	21,7	13,3	8,2	21,3	28,5	5,8	1,3
Niğde	15,1	5,2	7,5	45,1	23,9	2,2	1,1
Ordu	39,3	11,4	4,2	33,0	0,0	8,4	3,6
Rize	34,8	13,1	5,0	39,7	0,0	4,8	2,6
Sakarya	23,6	14,6	3,2	47,8	0,3	2,9	7,6
Samsun	32,6	11,7	5,0	41,0	2,3	3,9	3,5
Siirt	21,9	7,0	10,0	50,7	0,5	6,9	3,0
Sinop	45,9	14,7	6,1	20,3	1,1	8,3	3,7
Sivas	24,3	8,5	5,8	49,7	1,8	5,7	4,2
Tekirdağ	10,0	4,0	0,9	78,6	0,1	0,8	5,7
Tokat	37,3	11,9	5,0	30,0	1,7	6,0	8,1
Trabzon	44,2	16,9	5,4	18,9	0,1	6,1	8,4
Tunceli	38,4	9,9	27,1	4,7	0,1	9,7	10,0
Ş. urfa	30,2	8,9	3,4	17,8	23,0	2,8	13,8
Uşak	17,5	9,4	1,6	62,6	1,4	2,0	5,5
Van	41,0	16,9	9,4	12,4	0,1	6,6	13,6
Yozgat	29,3	10,6	3,5	28,9	11,0	7,2	9,5
Zonguldak	13,7	3,6	2,9	76,2	0,0	2,2	1,4
Aksaray	26,1	10,5	6,0	21,8	29,7	5,0	0,8
Bayburt	52,7	13,8	7,8	11,9	0,9	8,7	4,2
Karaman	18,8	7,0	5,4	33,1	30,8	2,9	2,0
Kirikkale	25,5	9,2	4,1	49,8	0,7	4,1	6,5
Batman	28,6	11,0	5,3	33,0	2,5	8,6	11,0
Şirnak	37,6	9,7	20,4	3,4	2,1	0,7	26,0
Bartin	41,2	13,7	13,5	26,6	0,1	5,0	0,0
Ardahan	50,7	17,4	16,0	0,4	0,0	9,2	6,3
Iğdir	47,8	14,2	18,4	6,7	0,0	10,4	2,5
Yalova	20,4	11,4	5,1	60,0	0,3	0,4	2,2
Karabük	19,6	5,5	2,4	69,3	0,1	2,3	0,7
Kilis	41,0	9,8	10,2	16,7	7,4	5,6	9,3
Osmaniye	27,7	10,3	6,2	47,3	2,0	1,8	4,7
Düzce	26,1	11,5	6,5	47,0	0,0	3,8	5,2
Total	**24,4**	**14,8**	**4,5**	**46,2**	**2,9**	**2,5**	**4,7**

As an example data set for one year, our data goes back for ten years for 81 cities. The subtitles under consumption include the following:

(i) House: the electricity usage in ordinary homes by ordinary citizens.

(ii) Commercial: the consumption in commercial buildings like supermarkets, small enterprises, trade centers, and so forth.

(iii) Governmental: the consumption in all governmental buildings and facilities like ministries, schools, courts, public parks and universities, and so forth.

(iv) Industry: energy consumption by industry representing the usage in industrial buildings as factories and any kind of production facilities.

(v) Agriculture: the consumption especially in fields or greenhouse facilities for irrigation, spraying, pruning or planting, and so forth.

(vi) Street lighting: the electricity consumption for lighting the streets and roads during nights.

(vii) Others: the amount of illegal usage of electricity.

B. Clustering Algorithm for MATLAB and Results

The Algorithm

```
[D,txt] = xlsread('inputs.xls');
%idx_som will give the group indices of each element for som clustering
[idx_som s] = som_clust(D, 'maxclust', 5, [10 10]);
idx_som
%IDX_kmeans will contain the group numbers of cities
%group centers will contain the group averages
[IDX_kmeans, group_centers] = KMEANS(D', 5);
figure;bar(hist(IDX_kmeans));title('k-meansclustering histogram');
IDX_kmeans
Cities_list = txt(3:end,1);
xlswrite('output_som.xls'cities_list, 81, 'A1:A81');
xlswrite('output_som.xls', idx_som, 81, 'B1:B81');
xlswrite('output_kmeans.xls', cities_list, 81, 'A1:A81');
xlswrite('output_kmeans.xls',
IDX_kmeans, 81, 'B1:B81');
```

Results

Cluster 1 Cities. "İstanbul"

Cluster 2 Cities. "Adana", "Denizli", "Gaziantep", "Hatay", "Kayseri", "Konya", "Manisa", "Mersin", "Tekirdağ", "Zonguldak", "Çanakkale".

Cluster 3 Cities. "Bursa", "Kocaeli", "İzmir"

Cluster 4 Cities. "Adıyaman", "Aksaray", "Amasya", "Artvin", "Aydın", "Balıkesir", "Batman", "Bayburt", "Bolu", "Burdur", "Bilecik", "Bingöl", "Bitlis", "Düzce", "Diyarbakır", "Edirne", "Elazığ", "Erzurum", "Erzincan", "Eskişehir", "Gümüşhane", "Giresun", "Isparta", "Iğdır", "Karabük", "Karaman", "Kars", "Kastamonu", "Kırıkkale", "Kırklareli", "Kırşehir", "Kütahya", "Kilis", "Malatya", "Mardin", "Muğla", "Muş", "Nevşehir", "Niğde", "Osmaniye", "Sakarya", "Samsun", "Sinop", "Sivas", "Siirt", "Tokat", "Trabzon", "Tunceli", "Uşak", "Van", "Yalova", "Yozgat", "Çankırı", "Çorum", "Şırnak"

Cluster 5 Cities. "Ankara", "Antalya"

C. Clustering via Rapid Miner and Results

See Table 3.

D. Clustering Results according to Solar Power Potential with Algorithm for MATLAB

See Table 4.

```
[num,txt,raw] = xlsread('SSAPMA.xls','SSAPMAS');
cities = txt(3:end,1);
[idx,c] = kmeans(num,5);
for kume = 1:5
disp('================='),disp(['Category ' num2str(cluster)])
disp('=================')
disp(sehirler(idx==cluster))
disp(['Center of cluster = ' num2str(c(cluster))])
end
```

Results

Category 1 (center of cluster = 2.54 (the median point of the values in the same cluster)). "Agri", "Ankara", "Aydin", "Balikesir", "Bartin", "Bilecik", "Bolu", "Burdur", "Bursa", "Cankiri", "Corum" "Denizli", "Duzce", "Erzincan", "Gaziantep", "Karabuk", "Kastamonu", "Kilis", "Kirikkale", "Kocaeli", "Kutahya", "Mugla", "Osmaniye", "Sakarya", "Usak", "Zonguldak"

Category 2 (center of cluster = 1.7922). "Ardahan", "Artvin", "Rize", "Trabzon"

Category 3 (center of cluster = 2.6791). "Adana", "Afyonkarahisar", "Bingol", "Canakkale", "Eskisehir", "Isparta", "Izmir", "Karaman", "Kirsehir" "Konya", "Manisa", "Nigde", "Sivas", "Yalova", "Yozgat"

Category 4 (center of cluster= 2.8492). "Adiyaman", "Aksaray", "Batman", "Bitlis", "Diyarbakir", "Edirne", "Elazig", "Istanbul", "Kahramanmaras", "Kayseri", "Kirklareli", "Malatya", "Mardin", "Mus", "Nevsehir", "Sanliurfa" "Siirt", "Sirnak", "Tekirdag" "Tunceli"

Category 5 (center of cluster = 2.27). "Amasya", "Antalya", "Bayburt", "Erzurum", "Giresun", "Gumushane", "Hakkari", "Hatay", "Igdir", "Kars", "Mersin", "Ordu", "Samsun", "Sinop", "Tokat", "Van"

E. Clustering Results according to Solar Power Potential with Rapid Miner (by *K*-means and SOM)

See Table 5.

F. Population and Population Forecasting of the Cities

See Table 6.

G. Welfare Indexes

See Table 7.

Table 3: Clustering of cities according to their consumption levels via Rapid Miner (by K-Means and SOM).

Cities	K-Means	SOM	Cities	K-Means	SOM	Cities	K-Means	SOM	Cities	K-Means	SOM	Cities	K-Means	SOM
Adana	5	5	Corum	3	5	Kayseri	5	5	Sinop	1	4	Burdur	3	5
Adıyaman	5	5	Denizli	3	5	Kilis	3	5	Sırnak	5	5	Bursa	3	5
A. karahisar	3	5	Diyarbakir	5	5	Kırıkkale	3	5	Sivas	3	5	Canakkale	5	5
Agrı	3	5	Duzce	3	4	Kırklareli	2	4	Tekirdag	2	5	Cankiri	3	1
Aksaray	5	5	Edirne	5	4	Kırsehir	5	5	Tokat	1	5	Karabuk	3	5
Amasya	1	5	Elazıg	5	5	Kocaeli	3	5	Trabzon	4	3	Karaman	3	5
Ankara	3	5	Erzincan	3	5	Konya	5	5	Tunceli	5	5	Kars	1	5
Antalya	1	5	Erzurum	1	5	Kutahya	3	5	Usak	3	5	Kastamonu	3	5
Ardahan	4	3	Eskisehir	3	5	Malatya	5	5	Van	1	2	Kars	1	5
Artvin	4	1	Gaziantep	3	1	Manisa	3	5	Yalova	3	5	Rize	4	2
Aydın	3	5	Giresun	1	5	Mardin	2	2	Yozgat	3	5	K. maras	5	5
Balıkesir	3	5	Gumushane	1	5	Mersin	1	5	Zonguldak	3	2	Bolu	3	5
Bartın	3	5	Hakkari	1	5	Mugla	3	5	Sakarya	3	5			
Batman	2	5	Hatay	1	5	Mus	5	5	Samsun	3	4			
Bayburt	1	5	Iğdır	1	5	Nevsehir	5	5	Sanlıurfa	5	5			
Bilecik	3	1	Isparta	3	5	Nigde	5	5	Siirt	5	5			
Bingol	5	5	İstanbul	5	5	Ordu	1	5	Karabuk	3	5			
Bitlis	5	1	İzmir	3	1	Osmaniye	3	5	Karaman	3	5			

TABLE 4

City	Sta. deviation (radiation time)	City	Sta. deviation (radiation time)
Adana	2,690296818	Kars	2,376578277
Adıyaman	2,850568535	Kastamonu	2,505402357
Afyonkarahısar	2,652801847	Kayserı	2,819897555
Agrı	2,531419233	Kılıs	2,441220522
Aksaray	2,793338678	Kırıkkale	2,590015551
Amasya	2,348400165	Kırklarelı	3,134691527
Ankara	2,575261583	Kırsehır	2,712404632
Antalya	2,362896038	Kocaelı	2,569262081
Ardahan	1,898816883	Konya	2,710960649
Artvın	1,667970115	Kutahya	2,54169071
Aydın	2,575892914	Malatya	2,835966791
Balıkesır	2,573129849	Manısa	2,631594758
Bartın	2,600760973	Mardın	2,976130971
Batman	2,902839127	Mersın	2,352446717
Bayburt	2,221639785	Mugla	2,474877522
Bılecık	2,559909938	Mus	2,781276186
Bıngol	2,74762966	Nevsehır	2,80581935
Bıtlıs	2,790823753	Nıgde	2,761498562
Bolu	2,513181361	Ordu	2,099196109
Burdur	2,562343762	Osmanıye	2,601296018
Bursa	2,56136506	Rıze	1,754007317
Canakkale	2,746524572	Sakarya	2,537168561
Cankırı	2,464731223	Samsun	2,156772763
Corum	2,476598249	Sanlıurfa	2,792549114
Denızlı	2,553234332	Sıırt	2,811756017
Dıyarbakır	2,883200864	Sınop	2,30354437
Duzce	2,539263208	Sırnak	2,823885719
Edırne	2,797553867	Sıvas	2,625525344
Elazıg	2,864982548	Tekırdag	2,950707425
Erzıncan	2,534056229	Tokat	2,370434061
Erzurum	2,375284194	Trabzon	1,848153733
Eskısehır	2,663166974	Tuncelı	2,79146219
Gazıantep	2,567747067	Usak	2,522644115
Gıresun	2,102217579	Van	2,347192768
Gumushane	2,256866682	Yalova	2,639856372
Hakkarı	2,140370392	Yozgat	2,673903072
Hatay	2,163409571	Zonguldak	2,56165135
Igdır	2,342966276	Kahramanmaras	2,790663377
Isparta	2,615078074	Karabuk	2,505135143
Istanbul	2,785876458	Karaman	2,639941471
Izmır	2,674965213		

TABLE 5: Clustering of cities according to their solar power potential via Rapid Miner (by K-Means and SOM).

Cities	K-Means	SOM	Cities	K-Means	SOM	Cities	K-Means	SOM	Cities	K-Means	SOM	Cities	K-Means	SOM
Adana	3	5	Corum	1	5	Kayseri	4	5	Sinop	5	5	Burdur	1	5
Adıyaman	4	5	Denizli	1	5	Kilis	1	5	Sırnak	4	5	Bursa	1	5
A. karahisar	3	5	Diyarbakir	4	5	Kırıkkkale	1	5	Sivas	3	5	Canakkale	3	5
Agrı	1	5	Duzce	1	5	Kırklareli	4	4	Tekirdag	4	5	Cankiri	1	5
Aksaray	4	4	Edirne	4	5	Kırsehir	3	5	Tokat	5	5	Karabuk	1	5
Amasya	5	5	Elazıg	4	5	Kocaeli	1	5	Trabzon	2	3	Karaman	3	5
Ankara	1	5	Erzincan	1	5	Konya	3	5	Tunceli	4	5	Kars	5	5
Antalya	5	5	Erzurum	5	5	Kutahya	1	5	Usak	1	5	Kastamonu	1	5
Ardahan	2	3	Eskisehir	3	5	Malatya	4	5	Van	5	5	Kars	5	5
Artvin	2	1	Gaziantep	1	5	Manisa	3	5	Yalova	3	5	Rize	2	2
Aydın	1	5	Giresun	5	5	Mardin	4	5	Yozgat	3	5	K. maras	4	5
Balıkesir	1	5	Gumushane	5	5	Mersin	5	5	Zonguldak	1	2	Bolu	1	5
Bartın	1	5	Hakkari	5	5	Mugla	1	5	Sakarya	1	5			
Batman	4	4	Hatay	5	4	Mus	4	5	Samsun	5	5			
Bayburt	5	5	Iğdır	5	5	Nevsehir	4	5	Sanlıurfa	4	5			
Bilecik	1	5	Isparta	3	5	Nigde	3	5	Siirt	4	5			
Bingol	3	5	İstanbul	4	5	Ordu	5	5	Karabuk	1	5			
Bitlis	4	5	İzmir	3	5	Osmaniye	1	5	Karaman	3	5			

TABLE 6: Population and population forecasting.

City	2010	2013	2018	City	2010	2013	2018
Adana	2.053.465	2.088.667	2.132.063	Manisa	1.342.281	1.379.025	1.435.691
Adıyaman	586.015	585.186	576.566	K. Maraş	1.070.647	1.131.705	1.230.314
Afyon	699.435	699.222	690.944	Mardin	761.087	775.180	792.768
Ağrı	529.263	522.044	502.477	Muğla	824.009	872.597	952.629
Amasya	321.808	317.933	308.935	Muş	386.845	358.135	303.076
Ankara	4.686.009	4.885.314	5.200.437	Nevşehir	284.083	287.204	291.079
Antalya	1.960.779	2.112.823	2.364.692	Niğde	350.013	366.502	391.436
Artvin	162.654	156.276	144.708	Ordu	721.843	723.376	720.102
Aydın	995.256	1.039.498	1.111.880	Rize	320.057	320.372	319.334
Balıkesir	1.151.826	1.182.196	1.227.914	Sakarya	874.881	908.937	962.535
Bilecik	193.516	193.554	192.431	Samsun	1.243.063	1.254.323	1.266.147
Bingöl	262.113	271.191	285.864	Siirt	317.155	344.970	395.875
Bitlis	323.902	317.820	303.224	Sinop	202.463	204.724	207.404
Bolu	268.277	266.087	259.305	Sivas	620.350	599.974	555.460
Burdur	246.624	244.404	238.345	Tekirdağ	834.868	932.255	1.097.284
Bursa	2.629.919	2.813.935	3.122.382	Tokat	619.250	619.802	613.781
Çanakkale	480.358	488.070	499.377	Trabzon	751.839	754.895	756.547
Çankırı	178.941	182.606	187.140	Tunceli	87.764	89.706	92.700
Çorum	536.874	521.868	491.257	Şanlıurfa	1.630.731	1.716.085	1.856.962
Denizli	934.733	958.709	994.970	Uşak	334.720	334.453	331.276
Diyarbakır	1.531.760	1.587.863	1.670.933	Van	1.051.872	1.125.180	1.250.414
Edirne	393.537	391.225	385.850	Yozgat	464.120	430.652	366.183
Elazığ	548.206	547.831	543.067	Zonguldak	620.069	620.198	616.773
Erzincan	210.123	208.845	205.292	Aksaray	378.316	388.990	404.248
Erzurum	744.843	694.706	597.331	Bayburt	72.809	68.106	59.193
Eskişehir	768.363	807.893	871.942	Karaman	236.899	246.889	263.218
Gaziantep	1.673.920	1.763.230	1.900.432	Kırıkkale	272.884	261.793	239.463
Giresun	425.732	430.608	436.138	Batman	506.690	539.309	595.643
Gümüşhane	133.805	137.039	141.200	Şırnak	458.158	504.989	591.400
Hakkâri	280.518	314.949	375.488	Bartın	190.357	197.658	209.303
Hatay	1.449.059	1.500.126	1.578.655	Ardahan	107.690	100.441	87.273
Isparta	412.916	419.760	427.768	Iğdır	182.881	180.291	173.562
Mersin	1.626.691	1.659.526	1.705.843	Yalova	220.737	256.136	315.925
İstanbul	13.050.933	13.571.135	14.407.233	Karabük	216.389	216.172	214.829
İzmir	3.900.770	4.055.605	4.304.649	Kilis	124.843	130.407	138.891
Kars	302.562	286.295	253.943	Osmaniye	480.250	518.737	541.319
Kastamonu	361.138	361.121	358.235	Düzce	337.201	349.642	369.165
Kayseri	1.217.110	1.264.676	1.339.823	Kütahya	556.534	539.844	505.554
Kırklareli	337.576	338.173	338.349	Malatya	739.902	748.731	761.343
Kırşehir	220.861	217.098	208.230				
Kocaeli	1.573.424	1.699.004	1.908.939				
Konya	2.004.743	2.050.921	2.111.845				

TABLE 7: Welfare indexes.

	First term (1990–1994)						Second term (1995–2010)				
Rank	City	Index value	Rank	City	Index value	Rank	City	Index value	Rank	City	Index value
1	İstanbul	4,270	41	Artvin	−,278	1	İstanbul	4,465	41	Kütahya	−,079
2	İzmir	2,689	42	Afyon	−,309	2	İzmir	2,549	42	Kastamonu	−,115
3	Ankara	2,441	43	Kastamonu	−,310	3	Ankara	2,477	43	Karaman	−,117
4	Bursa	1,577	44	Giresun	−,327	4	Bursa	1,651	44	Malatya	−,150
5	Kocaeli	1,574	45	Çorum	−,358	5	Kocaeli	1,355	45	Giresun	−,172
6	Adana	1,088	46	Sivas	−,434	6	Eskişehir	1,163	46	Niğde	−,177
7	Tekirdağ	1,070	47	Ordu	−,441	7	Adana	1,127	47	Afyon	−,183
8	Antalya	1,028	48	Niğde	−,450	8	Tekirdağ	1,077	48	Çorum	−,242
9	Eskişehir	,969	49	Aksaray	−,468	9	Antalya	1,068	49	K. Maraş	−,274
10	Balıkesir	,901	50	K. Maraş	−,469	10	Muğla	,915	50	Osmaniye	−,308
11	İçel	,863	51	Erzincan	−,473	11	Balıkesir	,912	51	Bartın	−,334
12	Aydın	,784	52	Çankırı	−,506	12	Kırklareli	,760	52	Sivas	−,396
13	Muğla	,742	53	Diyarbakır	−,532	13	Edirne	,721	53	Sinop	−,397
14	Kırklareli	,660	54	Sinop	−,540	14	Denizli	,719	54	Aksaray	−,402
15	Çanakkale	,606	55	Tokat	−,547	15	Çanakkale	,717	55	Çankırı	−,427
16	Konya	,590	56	Yozgat	−,656	16	Konya	,712	56	Tokat	−,439
17	Denizli	,561	57	Erzurum	−,679	17	Aydın	,644	57	Tunceli	−,439
18	Manisa	,559	58	Şanlıurfa	−,753	18	İçel	,599	58	Erzincan	−,468
19	Kayseri	,553	59	Adıyaman	−,849	19	Yalova	,580	59	Ordu	−,471
20	Edirne	,492	60	Tunceli	−,859	20	Manisa	,518	60	Yozgat	−,519
21	Bilecil	,404	61	Gümüşhane	−,904	21	Zonguldak	,483	61	Erzurum	−,639
22	Isparta	,394	62	Bayburt	−1,023	22	Kayseri	,397	62	Gümüşhane	−,709
23	Zonguldak	,338	63	Batman	−1,048	23	Hatay	,389	63	Diyarbakır	−,743
24	Sakarya	,337	64	Kars	−1,062	24	Sakarya	,371	64	Kilis	−,748
25	Hatay	,328	65	Siirt	−1,084	25	Bilecik	,368	65	Şanlıurfa	−,834
26	Bolu	,198	66	Mardin	−1,094	26	Bolu	,347	66	Adıyaman	−,919
27	Kırıkkale	,190	67	Van	−1,179	27	Samsun	,342	67	Bayburt	−1,017
28	Burdur	,168	68	Hakkâri	−1,242	28	Isparta	,329	68	Kars	−1,083
29	Samsun	,139	69	Bingöl	−1,293	29	Burdur	,295	69	Iğdır	−1,093
30	Uşak	,108	70	Bitlis	−1,340	30	Kırıkkale	,172	70	Ardahan	−1,157
31	Elazığ	,057	71	Şırnak	−1,356	31	Uşak	,159	71	Mardin	−1,163
32	Trabzon	−,004	72	Ağrı	−1,553	32	Trabzon	,142	72	Batman	−1,166
33	Nevşehir	−,006	73	Muş	−1,566	33	Rize	,111	73	Bingöl	−1,230
34	Rize	−,051				34	Karabük	,110	74	Siirt	−1,245
35	Kütahya	−,058				35	Nevşehir	,078	75	Van	−1,312
36	Kırşehir	−,096				36	Amasya	,038	76	Bitlis	−1,439
37	Malatya	−,104				37	Artvin	,013	77	Muş	−1,483
38	Gaziantep	−,118				38	Elazığ	,011	78	Şırnak	−1,542
39	Amasya	−,119				39	Gaziantep	,006	79	Ağrı	−1,465
40	Karaman	−,141				40	Kırşehir	−,034	80	Hakkâri	−1,659

Disclosure

The project "Saving industrial water with using solar power within the framework of artificial intelligence" that was prepared by the corresponding author via utilizing the same methods about water management won the winner prize in UNDP-Every Drops Matter contest in 2014 [43].

Competing Interests

The authors declare that they have no competing interests.

References

[1] *World Economic Outlook*, International Monetary Fund, 2011.

[2] R. Florida and M. Atlas, "Report of field research on environmentally conscious manufacturing," *California Management Review*, vol. 39, no. 2, pp. 80–105, 1997.

[3] M. Grubb, *Renewable Energy Strategies for Europe V1-Foundations and Context*, The Royal Institute of International Affairs, London, UK, 1995.

[4] B. Pepukaye, A Framework for Sustainable Strategic Energy Company Investment Analysis and Decision-support: Towards the Operationalization of Sustainable Energy Systems-Lund, Sweden, 2001, http://lup.lub.lu.se/record/20261.

[5] U. Deichmann and F. Zhang, *Growing Green*, World Bank, 2013.

[6] J. Jackson, "Data mining: a conceptual overview," *Communication of the Association for Information Systems*, vol. 8, pp. 267–296, 2002.

[7] International Renewable Energy Agency (IRENA), *Rethinking Energy: Renewable Energy and Climate Change*, 2014, http://www.irena.org/rethinking/IRENA_REthinking_Energy_Presentation_2014.pdf.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2001.

[9] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT Press, 2001.

[10] H. M. Chung and P. Gray, "Special section: data mining," *Journal of Management Information Systems*, vol. 16, no. 1, p. 11, 1999.

[11] T. W. Liao and E. Triantaphyllou, Eds., *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*, World Scientific, Singapore, 2007.

[12] W. Zhang, *Computational Ecology-Artificial Neural Networks and Their Applications*, World Scientific, Singapore, 2010.

[13] M. Minkov and G. Hofstede, "Clustering of 316 european regions on measures of values: do Europe's Countries Have National Cultures?" *Cross-Cultural Research*, vol. 48, no. 2, pp. 144–176, 2014.

[14] P. Seidler and R. Adderley, "Criminal network analysis inside law enforcement agencies: a data mining system approach under the National Intelligence Model," *International Journal of Polica Science and Management*, vol. 15, no. 4, pp. 323–337, 2013.

[15] M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie, and J. S. Brownstein, "Combining search, social media, and traditional data sources to improve influenza surveillance," *PLoS Computational Biology*, vol. 11, no. 10, Article ID e1004513, 2015.

[16] P. Sobkowicz, M. Kaschesky, and G. Bouchard, "Opinion mining in social media: modeling, simulating, and forecasting political opinions in the web," *Government Information Quarterly*, vol. 29, no. 4, pp. 470–479, 2012.

[17] M. Esien, *Cluster and Tree View Manual*, Stanford University, 1998.

[18] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[19] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, "Engineering applications of the self-organizing map," *Proceedings of the IEEE*, vol. 84, no. 10, pp. 1358–1384, 1996.

[20] T. Kohonen and S. Kaski, *Exploratory Data Analysis by the Self Organizing Maps: Structure of Welfare and Poverty in the World*, Helsinki University of Technology, Singapore, 1996.

[21] T. Kohonen and S. Kaski, *Exploratory Data Analysis by The Self Organizing Maps: Structure of Welfare and Poverty in the World*, Helsinki University of Technology, Singapore, 1996.

[22] S. Kaski, *Data Exploration Using Self-Organizing Maps*, Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series no. 82, Finnish Academy of Technology, Espoo, Finland, 1997.

[23] S. A. Mingoti and J. O. Lima, "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1742–1759, 2006.

[24] S. J. Fraser and B. L. Dickson, *Data Mining Geoscientific Data Sets Using Self Organizing Maps*, 2005.

[25] Y. Miche, A. Akusok, D. Veganzones et al., "SOM-ELM—self-organized clustering using ELM," *Neurocomputing*, vol. 165, pp. 238–254, 2015.

[26] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*, Michigan State University, Michigan, NJ, USA, 1988.

[27] P. Bradley and U. Fayyad, "Refining initial points for K-means clustering," in *Proceedings of the International Conference on Machine Learning (ICML '98)*, pp. 91–99, January 1998.

[28] E. Alpaydın, *Introduction to Machine Learning*, MIT Press, Boston, Mass, USA, 2004.

[29] S. Chen, X. Yang, and Y. Tian, "Discriminative hierarchical K-means tree for large-scale image classification," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 26, no. 9, pp. 2200–2205, 2015.

[30] N. Elssied, O. Ibrahim, and A. Osman, "Enhancement of spam detection mechanism based on hybrid k-mean clustering and support vector machine," *Methodologies & Applications*, vol. 19, no. 11, pp. 3237–3248, 2015.

[31] D. A. Zighed, S. Tsumoto, Z. W. Ras, and H. Hacid, Eds., *Mining Complex Data*, vol. 165 of *Studies in Computational Intelligence*, Springer, New York, NY, USA, 2009.

[32] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, NJ, USA, 5th edition, 2002.

[33] L. Borges, V. Marques, and J. Bernardino, "Comparison of data mining techniques and tools for data classification," in *Proceedings of the ACM International Conference on Computer Science and Software Engineering (C3S2E '13)*, pp. 113–116, Porto, Portugal, July 2013.

[34] A. Jović, K. Brkić, and N. Bogunović, "An overview of free software tools for general data mining," in *Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO '14)*, pp. 1112–1117, Opatija, Croatia, May 2014.

[35] G. Piatetsky, *KDnuggets Annual Software Poll: RapidMiner and R vie for First Place*, 2013, http://www.kdnuggets.com/2013/06/kdnuggets-annual-software-poll-rapidminer-r-vie-for-first-place.html.

[36] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, University of Illinois at Urbana-Champaign, Elsevier, 2nd edition, 2016.

[37] Energy Trilemma Index and World Energy Council, *Project Partner Wyman O.*, 2015.

[38] P. Avato and J. Coony, *Accelerating Clean Energy Technology Research*, Development and Deployment, The International Bank for Reconstruction and Development-The World Bank, 2008.

[39] O. A. Abbas, "Comparisons between data clustering algorithms," *The International Arab Journal of Information Technology*, vol. 5, no. 3, pp. 320–325, 2008.

[40] S. S. Shahapurkar and M. K. Sundareshan, "Comparison of self-organizing map with K-means hierarchical clustering for bioinformatics applications," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 1221–1226, July 2004.

[41] A. de Castro Leão, A. Duarte Dória Neto, and M. Bernardete Cordeiro de Sousa, "New developmental stages for common marmosets (*Callithrix jacchus*) using mass and age variables obtained by K-means algorithm and self-organizing maps (SOM)," *Computers in Biology and Medicine*, vol. 39, no. 10, pp. 853–859, 2009.

[42] F. Bação, V. Victor Lobo, and M. Painho, "Self-organizing maps as substitutes for K-means clustering," in *Computational Science—ICCS 2005: 5th International Conference, Atlanta, GA, USA, May 22–25, 2005, Proceedings, Part III*, vol. 3516 of *Lecture Notes in Computer Science*, pp. 476–483, Springer, Berlin, Germany, 2005.

[43] http://www.everydropmatters.com/every-drop-matters-water-awards-winners/.