DOI 10.33111/nfmte.2021.086

EU COUNTRIES CLUSTERING FOR THE STATE OF FOOD SECURITY USING MACHINE LEARNING TECHNIQUES

Vitaliy Kobets

Kherson State University
27 Universitetska Str., Kherson, 73003, Ukraine
ORCID: 0000-0002-4386-4103, E-mail: vkobets@kse.org.ua

Oleksandra Novak

Institute of International Relations,
Taras Shevchenko National University of Kyiv
36/1 Yuriia Illienka Str., Kyiv, 04119, Ukraine
ORCID: 0000-0002-5703-7416, E-mail: olexandranovak@gmail.com

The food security problem has emerged from the growing pressure of demographic problem and global inequality. Overall, the state of food security is optimal in the EU. This was achieved due to effective implementation of regulatory initiatives regarding EU countries food self-sufficiency and intra-EU food market protection. The purpose of the research paper was to cluster EU countries in terms of food security level using advanced mathematical modeling tools. To this end, we selected 5 food security factors (FAO Food production index, Total factor productivity in agriculture, Per capita agricultural expenditure, Consumer prices food, Net trade food index) to which we applied the following cluster analysis algorithms (self-organizing maps, dendrograms, k-means and k-medoids clustering). As a result of the conducted experimental research, it was found that self-organizing maps and dendrograms methods to be better suited for data visualization, whereas k-means and k-medoids give more accurate and detailed solutions. The obtained results gave us an opportunity to define the advantages and disadvantages of the selected clustering methods, as well as to present agripolicy recommendations for different groups of EU countries.

Keywords: *cluster analysis, food security, self-organizing map, hierarchical clustering, k-means, k-medoids*

JEL Classification: C38, C88, L66, Q18

Introduction

The food security problem has been fundamentally considered since 1970s. The concept of “food security” was first used in 1974 at the World Food Conference in Rome. In 1975, the International Food

Policy Research Institute (IFPRI) was founded in Washington. During 1980s, Nobel laureate A. Sen regarded food security as a problem of households purchasing power that is affected by access to income and other resources (e.g. transfers and gifts), market integration, pricing and market conditions [1].

Proponents of the theory of "food regimes" (food regime theory), developed by H. Friedman and P. McMichael in the late 1980s [2], argue that there is a direct link between the development of food systems and periods of capital accumulation. Now this theory is used to explain the strategic role of agriculture and food in the context of the global food system.

Summarizing the approaches of the Food and Agriculture Organization (Rome Declaration, 1996) [3], the Committee on World Food Security (2012) [4] and the International Finance Corporation (2012) [5], we can define food security as the level of provision of essential foods from one's own source of income and their accessibility to all, in such volumes and assortment that best meet necessary and useful consumer needs [6].

The idea of "food sovereignty", proposed by the Via Campesina international farmers' association at the 1996 World Food Summit, is becoming more and more popular as an alternative to the existing neoliberal approach to the food problem. H. Wittman believes that "food sovereignty" requires such a model of trade relations that best meets the social, economic, political and environmental principles of the alternative food paradigm [7].

S. Maxwell discusses three changes in the paradigm of scientific thinking about food security in the early 2000s: first, the transition "from global and national to household and individual"; secondly, the transition from the views of "food first" to "livelihood perspective"; third, the change "from objective indicators to subjective perception" [8].

First of all, speaking about the trends in food security in the world, we should pay attention to the agricultural policy of individual states. The success of agricultural sector, therefore, depends not only on the level of technical equipment, as is with industry, but also on the ability of the producer or exporter of agricultural products to receive financial and administrative support. The high degree of public participation is explained by the strategic goals of ensuring national food security, supporting the development of agricultural areas, providing employment, solving social problems. OECD experts note that the level of protectionism in agricultural products is four times higher than the level of state support for non-agricultural products, so we can talk about the significant prevalence of agriprotectionism.

Ye. Novikov proposed to divide all countries in the world into 3 groups regarding on the attitude to food security problem, which is primarily taken into account in the development of agricultural policies [9], as shown in Fig. 1.

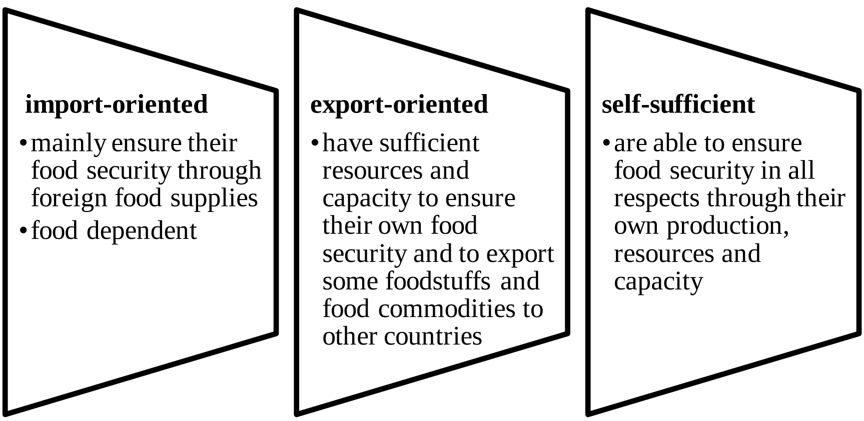


Fig. 1. Countries classification regarding food security problem

The issue of food security in the priority of EU agricultural policy. Primarily, EU initiatives focused on the formation of a common market based on the free movement of agricultural products between Member States. However, this was complicated by significant discrepancies in the sanitary and veterinary legislation of the Member States, which led to the emergence of non-tariff barriers to the flow of food and food commodities.

Nowadays food sovereignty in the EU means building a new model of agriculture that prohibits the use of industrial processes in food production, provides quality food for people, restores European farms that have disappeared due to competition from agricultural holdings and adjusts existing trade imbalances (arising from the concentration of agricultural production in certain geographical areas). This concept stipulates that state subsidies should not support the expansion of agribusiness in a geographical area that has a detrimental effect on the economy and social structure of others. An important strategic innovation that solves these problems is the agrarian paradigm of the bioeconomy, which was introduced by the EU, in particular by Germany in recent years, and later adopted by other countries [10].

The EU Common Agricultural Policy (CAP) is a program upon which the EU has been allocating money to support agriculture since 1957. The primary CAP goal was to achieve food self-sufficiency and to restrict food imports. Thanks to CAP measures, since 2010 the EU has transformed from a net food importer to the second (after the US) largest food exporter in the world as of 2019 [11] Therefore, the state of food security in the EU countries up to date depends on CAP mechanisms.

The topicality stems from the growing pressure of demographic problem and global inequality that do not contribute to solving the food problem. This has an equally negative impact on the living standards of all people either from the developing or the developed countries.

Various methods of mathematical modeling and forecasting were used to analyze food security. For example, a study of basic agricultural macro-models was presented in the paper [12]. The authors compare the models used to analyze the agricultural sector, medium- and long-term forecasts, and policy making. It concludes that a promising direction for modelling the agricultural economy and policy is the systematic application of stochastic analysis and risk management approaches.

In [13] the authors discuss the impact of COVID-19 on agriculture. In particular, the authors note that the pandemic has had a significant impact on food security and has identified weaknesses in agricultural policies. However, COVID-19 has opened new opportunities for the Ukrainian agricultural sector: acceleration and expansion of the digitalization process, the possibility of expanding the market for export products and an increase in skilled and unskilled labor.

Other authors [14] applied ML as well as traditional econometric methods in estimation of food security of householders on micro level. Since predictive accuracy depends partly on which indicators are used to identify food insecure households, it is important to assess the performance of calorie-based indicators. They found that overall prediction accuracy ranged from 60% to 70% for their chosen machine learning and other methods. They found that the ML and non-ML methods showed similar accuracy. Which methods will work best will depend on the pool of available predictor sets, as well as the complexity of the functional forms that link predictors to outcomes [14].

Authors [15] examine the relationship between food sharing and deprivation generally, before applying ML techniques to develop a predictive model of food insecurity based on aggregated food sharing behaviors. ML is driving this transformation through predictive and descriptive analytics (e.g., more sophisticated segmentation and summarization).

In the paper [16] authors develop tools for ensuring the economic security of agrarian sector on the example of Ukraine, which consists of the following units: information and analytical unit, regulatory influence unit, unit of counteraction to threats and control unit. To assess economic security, indicators of gross domestic product and labor productivity in the agricultural sector are used.

To the best of our knowledge, there are no machine learning models for the revealing of countries clusters with different levels of food security to identify adequate strategies of countries to comply with food security, depending on the initial conditions of these countries. So, the scientific novelty of our study lies in the formation of a unique set of factors for analyzing the state of food security in the European Union based on open statistics of the World Bank, WTO, Eurostat and FAO. Also, the advantages and disadvantages of the selected clustering methods were identified, and agripolicy recommendations for different EU countries clusters were presented.

The **purpose** of the research paper was to cluster EU countries in terms of food security level using advanced mathematical modeling tools and generate the recommendations on agricultural policy for various countries taking into account their initial conditions.

We organize the structure of our paper as follows: in introduction, we described the problem, made a critical review of the literature and stated the purpose of the study; in section 1, we propose research methodology; in section 2, we present clustering methods; in section 3, we justify the choice of food security factors selected for clusterization; section 4 is dedicated to interpreting the results of cluster analysis; finally, we present our conclusions.

Main body

1. Research Methodology

For the purposes of our research, it is relevant to carry out a cluster analysis to divide the EU countries into relatively homogeneous groups according to the level of food security. This implies that the countries (objects) within the group must be similar (homogeneous) in relation to each other and dissimilar (heterogeneous) to the countries (objects) of other groups in relevant factors. In other words, the greater the similarity (homogeneity) within one group is, the greater the difference between the groups is. Thus, the better the solution provided by cluster analysis is [17].

The final method of grouping clustering objects with a given cluster method significantly depends on the choice of the method of measuring the distance (determining the degree of similarity of objects). The similarity or difference between the classified objects is established depending on the metric distance between them (metrics). If each object is described by k features, it can be represented as a point in k-dimensional space, and the similarity with other objects will be defined as the corresponding distance. The distance between two objects is denoted as $d(x_i, y_i)$ – it is a nonnegative function of proximity set under the following conditions [18]:

1) always > or = 0;

2) distance from point X to point Y equals the distance from Y to X;

3) if the numerical values of the factors of the two objects are the same, the distance between them equals 0;

4) let there be a third point U, then the sum of the distances between the points XU and YU always > the distance between the points XY [19].

Given the features, advantages and disadvantages of different metrics, for the purposes of grouping EU countries by indicators of food security, for the purposes of our research it is advisable to use the quadratic Euclidean distance. Euclidean metric (d_E) is the most common function of the distance between two objects (x; y) and is formulated as follows: $d_E(x; y) = \sqrt{\sum_{i=1}^{Nf}(x_i - y_i)^2}$, where i – factor number, characterizing the object.

In general, Euclidean distance allows to omit the sign differences with a proportionally increasing distance between objects when absolute values of indicators are totally different [19]. This increases the dimensionality of the cluster field – objects are artificially separated from each other and as a result, the boundaries between clusters become clearer and more precise.

We prepared comparative characteristics of software for cluster analysis (Table 1). To begin with, we can use free software (Sciencehunter or NCSS Statistical software) to perform standard cluster analysis. Should we perform more detailed analysis, it is necessary to install libraries and code in R [20] or write own macros [21].

Table 1

COMPARATIVE CHARACTERISTICS OF APPLIED SOFTWARE FOR CLUSTER ANALYSIS

Software	**Sciencehunter**	**NCSS Statistical software**	**RStudio**	**Excel Macros**
Data analysis	automated mode	automated mode	using R code	using Visual Basic
Visualization and construction of clusters	automated mode	automated mode	manual operation	manual operation
Hierarchical and non-hierarchical algorithms	automated mode	automated mode	using R code	None
Data export	-	+	+	-
Clustering report	+/-	+	+	-
Fee	free	30 days trial	free	License of MS Office

To date, there are many algorithms for data clustering. In general, they are all divided into hierarchical and non-hierarchical (iterative). For the purposes of our research, we primarily performed hierarchical SOM (Self-Organizing Map) algorithm. Then we applied the most common cluster analysis techniques – dendrograms (Hierarchical Clustering/Dendrograms) and non-hierarchical k-means method (k-means Clustering). Since there was a need to present more accurate and comprehensive results, we used one more non-hierarchical k-medoids (Medoid Partioning) method. Below we will briefly describe methods' algorithms.

2. Clustering Methods

Fitting unsupervised self-organizing maps (SOM) is an iterative adjustment of the weight vector $\overline{\mathrm{w}}_j^t = \{w_{1j}^t, \dots, w_{ij}^t, \dots, w_{mj}^t\}$, consisting of the weights of all parameters $i = 0,1, \dots, m$ of each neuron $j = 1,2, \dots, p$, t – number of iteration. To this end, a modified Hebb competitive learning algorithm is used. It takes into account not only the score of the winning neuron, but also its nearest neighbors located in the *R*-proximity:

1. At the initialization stage small random values are assigned for all weights w_{ij}^0, $i = 0,1, \dots, m$.

2. The parameter vector of each object $\overline{\mathrm{y}} = \{y_1, \dots, y_i, \dots, y_m\}$ of the input layer are randomly fed to the network outputs and a "winning neuron" (BMU, Best Matching Unit) with a minimum distance is selected for each of them $\sum_{i=1}^{m}(y_i - w_{ij}^t)^2$ $\forall j = 1,2, \dots, p$.

3. The subset of the "immediate environment" of BMU is determined. And BMU's radius R decreases with each iteration t.

4. The weights $\overline{\mathrm{w}}_j^t$ of selected nodes are recomputed with a view to their distances to the winning neuron and their proximity to the vector $\overline{\mathrm{y}}$.

The method of dendrogram constructing (Hierachical Clustering / Dendrograms) is an agglomerative hierarchical clustering algorithm, which builds a hierarchy of clusters in the form of a tree diagram (dendrogram). In the beginning, each object forms a separate cluster, and in the process of constructing at each step, the two most similar clusters are combined into a single new cluster. After merging, objects are never separated [22]. There are 8 main linkage types in hierarchical analysis. In our paper we will use the most common type of Group Average (Unweighted Pair-Group).

The k-means method (k-means Clustering) developed by J. MacQueen in 1967 remains the most widely used method of following iteration procedures [23]. The k-means algorithm was described by J. Hartigan and M. Wong in 1979 as a partitioning technique [24] and is incorporated in NCSS program service [25]. Some researchers note that Hartigan's methodology is more accurate, as far as it is not inasmuch influenced by random primary location of the centroid. It is most useful for forming a small number of clusters from a large number of observations. It requires variables that are continuous with no outliers. Discrete (logic) data can be included but may cause problems. The objective of this technique is to divide *N* observations with *P* dimensions (variables) into *K* clusters so that the within-cluster sum of squares is minimized. The researcher itself chooses the number of clusters to be established. Since the number of possible arrangements is enormous, it is not practical to expect the best solution. Rather, this algorithm finds a "local" optimum (minimization of sums of squares of distances between each observation and its cluster center). There are so many iterations with different initial configurations until the centers of the clusters become stable (that is, each iteration has the same objects in each cluster). Then the most optimal of the obtained cluster solutions is selected (when the variance within the cluster will be minimized, and between the clusters – maximized).

The k-medoids method (Medoid Partioning) was first presented by H. Spath in 1985 as a method of minimizing an objective function by swapping objects from one cluster to another [26]. Beginning at a random starting configuration, the algorithm proceeds to a local minimum by intelligently moving objects from one cluster to another. When no object moving would result in a reduction of the objective function, the procedure terminates. Unfortunately, this local minimum is not necessarily the global minimum. To overcome this limitation, the program lets you rerun the algorithm using several random starting configurations and the best solution is saved. Contrary to k-means method, in k-medoids method data set points are chosen as a center (a medoid or an example).

3. Rationale for the Choice of Factors to Study the State of Food Security in the EU

The choice of factors on the basis of which the research will be carried out is of great importance for cluster analysis. It should be noted that after conducting an analytical review of food security in the EU, we came to the conclusion that there is no crisis situation in the region regarding the problem of hunger and malnutrition.

Therefore, we decided not to use traditional Food and Agriculture Organization (FAO) food safety indicators in terms of human health indicators (number of kcal consumed per person, protein consumed, number of women suffering from anemia, etc.) [27]. We tried to incorporate the factors that would be subordinated to the main drivers of global food security (demand, supply, international food trade), as shown in Fig. 2 [28].

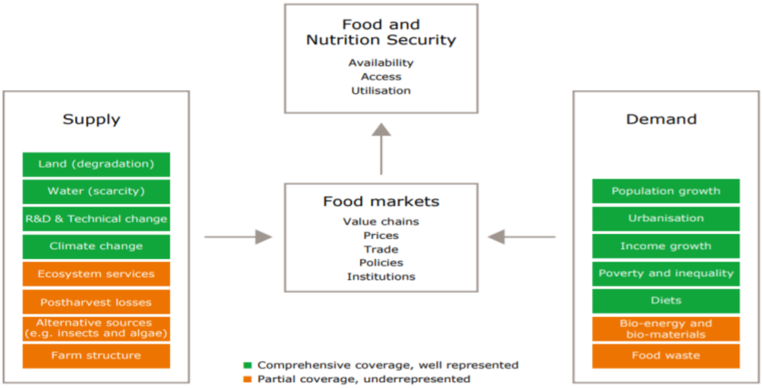


Fig. 2. Global drivers of food security

Also, the development of agro-industrial complex, state support of agro-industrial complex within the framework of CAP and international food trade significantly affect the state of food security in the EU countries. Therefore, we have chosen 5 following factors (Fig. 3):

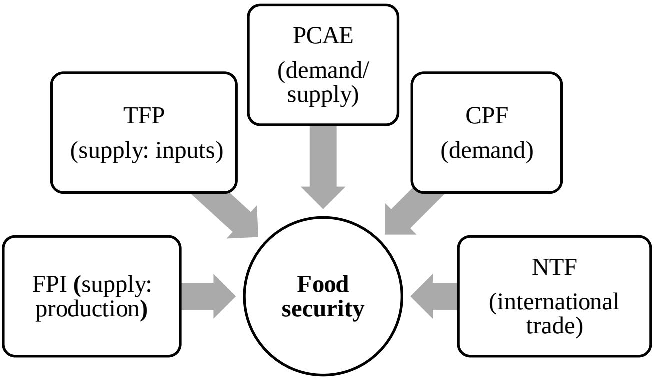


Fig. 3. Factors of cluster analysis

1) *FAO Food production index (FPI)* of the World Bank database. The FPI is calculated according to the Laspeyres index formula and estimates the relative level of total food production for consumption (production for seed or animal feed is not included in the index) for each year compared to the base period of 2004-2006 [29]. This indicator falls under the criterion of physical availability of food and is a supply factor.

2) *Total factor productivity in agriculture (TFP)* of Eurostat database being one of the contextual CAP indicators [30]. The TFP compares the total output relative to the total costs used in the production of agricultural products. As output and costs are expressed in terms of volume indices, the indicator reflects the growth rate of TFP. This indicator is a consolidated indicator of growth in productivity of land, capital and labour. The base year for TFP is 2005, and then a three-year flowing average is calculated to smooth out the effects of external factors, such as weather conditions. Thus, the data for 2016 correspond to the average index for (2015-2017). In recent years, the growth rate of TFP in the EU has slowed. The growth of TFP reflects the impact of technologies and innovations that increase the efficiency of the use of inputs of labour and capital in the production process [31]. This indicator falls under the criterion of sustainability of food systems and is a supply factor.

3) *Per capita agricultural expenditure (PCAE)*. This indicator has been calculated by the indicator of per capita agriculture support expenditures of IFPRI database that is a part of SPEED (Statistics on Public Expenditures for Economic Development) indicator system [32]. Actually, the expenditure indicator is expressed in constant prices ($) in 2011. However, for the purpose of our research, we decided to calculate the growth rate, taking for the base period the available data for 2005, and for the current period – data for 2016 by the formula:

$$T_{gr} = \frac{I_C}{I_b} \cdot 100\%,$$

where I_C – current indicator, I_b – base indicator.

The obtained PCAE indicator falls under the criterion of sustainability of food systems and acts as a supply factor. Its calculated values can be seen in Table A1 posted on Google Drive

[33]. However, given that the initial calculated data include the population in each country, we believe that this indicator is also a demand factor.

4) *Consumer prices food (CPF)* of FAOSTAT database [27]. This indicator has been calculated manually, as the database includes only monthly indices (base period – 2005), that show changes in per cent as compared to the corresponding month of the previous year. We used the summary table to obtain the average annual index for 2016 for each country, which is presented in Table A2 [33]. This indicator falls under the criterion of economic affordability of food and is a demand factor.

5) *Net trade food index (NTF)* has been calculated upon the dataset of WITS (World Integrated Trade Solution) database, shown in Table A3 [33]. In general, the Net trade food index shows for each commodity the level of exports excess over imports (with a positive value) or the level of imports excess over exports (with a negative value). The index is expressed in the range from -1 to +1 by the formula: $NTF = \frac{E_i - I_i}{E_i + I_i}$.

Extreme value "–1" means that the goods are only imported, i.e. the country only imports food, no exports. An extreme value of "+1" means that the product is only exported and there is no import. Of course, this is not feasible in the real world, so other negative values will show the degree of imports excess over exports (the country is a net food importer), and other positive values show the degree of exports excess over imports (the country is a net food exporter). Thus, to calculate this indicator, we took data on food exports and imports among EU countries from WITS database for 2016.

Since not all factors have open statistics for recent years, we took data for 2016 in our paper for raw data. The summary table with the calculated 5 factors of the cluster analysis is presented in Table A4 [33].

4. Cluster Analysis Results Interpretation

The basic software of our research are R (for SOM) and NCSS. The dataset consists of 5 food security factors expressed in index form. Therefore, before starting the cluster analysis procedure, the numerical data of the factors were standardized, as seen in Table A.5 [33].

To begin with, we constructed unsupervised self-organizing maps (Kohonen maps) using language R in RStudio framework. 5 factors (*FPI*, *TFP*, *PCAE*, *CPF* and *NTF*) described above are fed into an input layer of 5 neurons. The structure of the map 3×5 neurons allows to make better detailing of the obtained results. The efficiency of such a neural network structure has been experimentally confirmed.

We repeated steps 2-4 of the algorithm using function SOM of kohonen library in RStudio until the original values of the network stabilize with a given accuracy according to parameter alpha – monotonically decreasing learning rate. Be default, the value decreases linearly from 0.05 to 0.01. Thus, all 27 EU countries are self-organizing on the output layer neurons.

The average distance to the nearest neurons after 100 iterations is almost halved (Fig. 4).

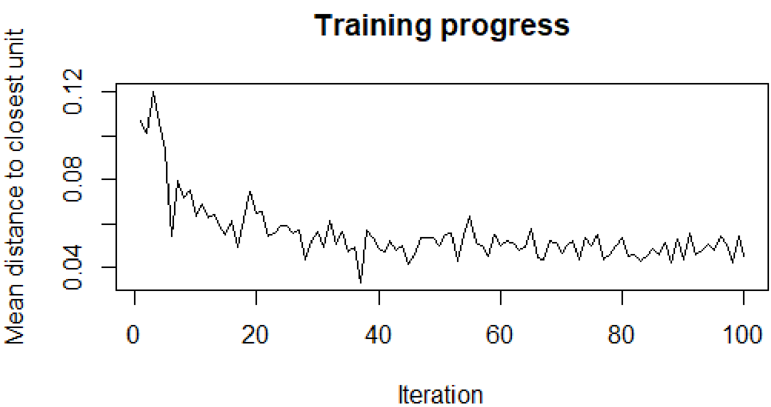


Fig. 4. Decrease in average distance to the nearest neurons after 100 learning iterations of the SOM network with a change in the values of the alpha parameter

The “codes” plot visually displays the value of 5 factors for each node, which corresponds to 27 countries (Fig. 5).

The *map$unit.classif* command allows to determine the correspondence of nodes and countries: {2 9 14 6 15 2 4 6 1 4 2 5 4 13 4 10 11 3 12 7 11 2 8 8 3 11 3}.

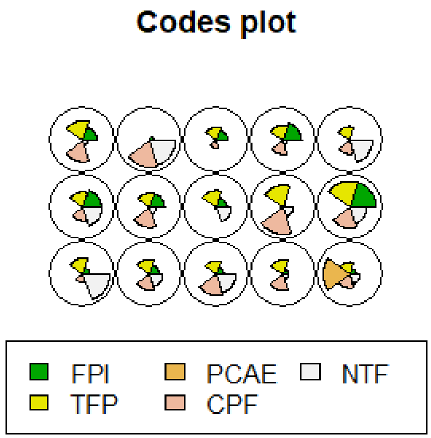


Fig. 5. SOM maps with "codes" type

Thus, the first country corresponds to node 2, the second country to node 9, the third country to node 14, and so on.

Map types "counts" and "mapping" allow to distribute countries on 15 nodes (the largest number in the 2nd, 3rd, 4th and 11th nodes from left to right from the bottom row of nodes to the top), as shown in Fig. 6. "Counts" graph on the left shows the distribution of countries by nodes by color, ranging from red (that indicates the smallest number of countries) to light yellow (that shows the largest number); "mapping" graph on the right indicates the distribution of countries by nodes where each country is denoted by a small circle).

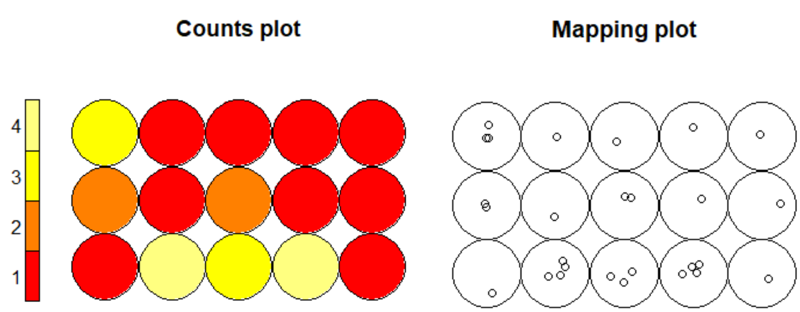


Fig. 6. SOM maps with "counts" and "mapping" types

The distance between country nodes is indicated at the map where the greater the distance is, the lighter the color is assigned and vice versa (Fig. 7). Red color signifies the smallest distance between country nodes, white – indicates the greatest distance, intermediate colors (from orange to yellow) show an increase in the distance between country nodes.

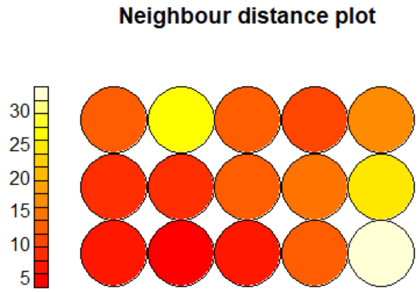


Fig. 7. SOM maps with "neighbor distance" type

For the number of clusters $k = 3$, we have performed hierarchical clustering via SOM algorithm and have constructed the maps of the "codes" type (with the distribution of the share of variables score). The results obtained are presented in Fig. 8.

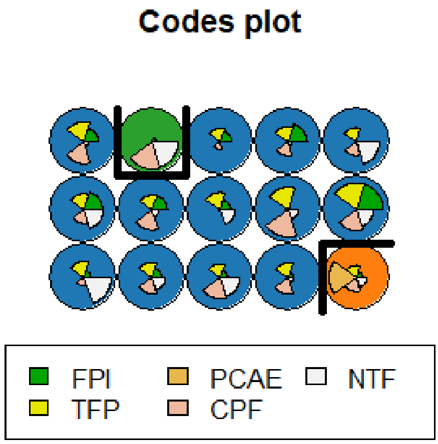


Fig. 8. Clustering of SOM map nodes

In the end, we see from the node/country ratio that Malta forms cluster 1, cluster 2 includes Greece and cluster 3 encompasses the rest of the countries. Characteristics of each cluster are set out in a third column of Table 2.

Table 2

CLUSTERS' COUNTRIES CHARACTERISTICS

Clusters	Countries	Characteristics
1 (green)	Malta	largest *CPF*, large *NTF*, no *PCAE*
2 (orange)	Greece	largest *PCAE*, less than average *TFP*
3 (blue)	25 other countries	importance of *TFP*

First cluster has highest consumer prices food index, net trade food (export substantially more than import) and almost has no state support (per capita agricultural expenditure).

Second cluster has very large state support, but smaller total factor productivity in agriculture.

Third cluster includes 25 countries, where the more consumer prices food the larger total factor productivity in agriculture, as can be seen from Table 2.

Since the obtained results include 2 clusters that are formed only with one country, we do not consider such distribution successful. Thus, we decided to apply a number of clustering methods in NCSS program.

In NCSS we started from applying hierarchical classification algorithm with the Group Average linkage type.

The dendrogram showed that Greece has formed a separate cluster, and the division of countries into other groups remained unclear (Fig. 9).

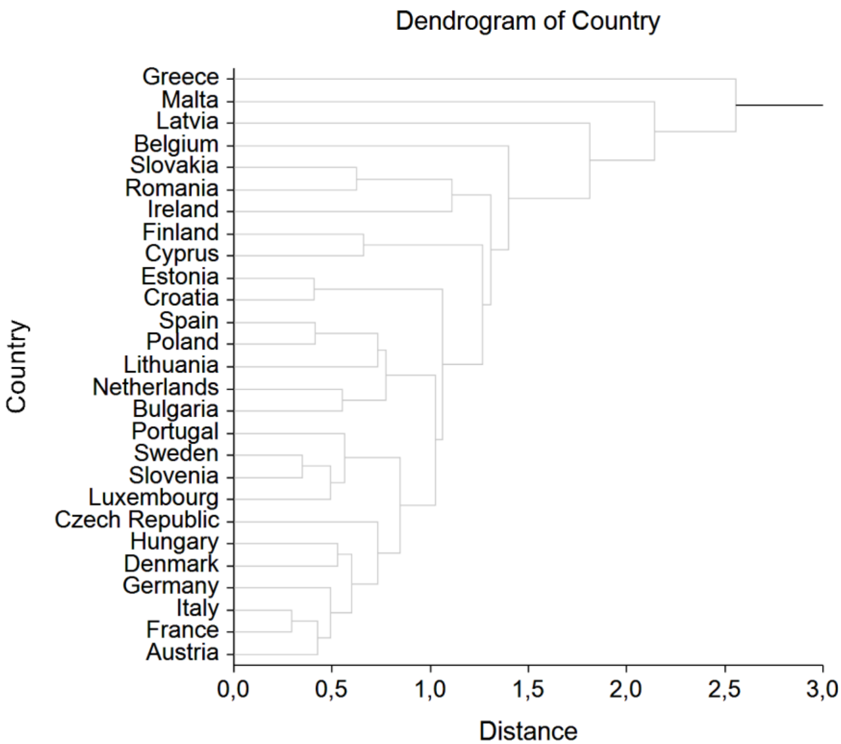


Fig. 9. The results of hierarchical cluster analysis in NCSS

If we set the division of the EU countries into 3 clusters, then Malta will also separate into a new cluster, as can be seen from Fig. 9. With an increase in the number of clusters, new clusters will be separately occupied by Latvia, Belgium, etc.

Therefore, we continued our research by applying k-means clustering and began with the allocation of 2 clusters that demonstrated the same problem as in the hierarchical analysis – Greece formed an independent cluster (Figs. 10, 11). Thus, Fig. 10 illustrates the k-means cluster analysis report in NCSS, and Fig. 11 shows two plots of clustering along the axes *FPI* and *PCAE*, and also *TFP* and *NTF*, respectively.

Variables	Cluster1	Cluster2
FPI	0,02689862	-0,6993641
TFP	0,03025652	-0,7866695
PCAE	-0,1935805	5,033094
CPF	0,01507262	-0,391888
NTF	0,01115227	-0,289959
Count	26	1

F-Ratio Section

Variables	DF1	DF2	Between Mean Square	Within Mean Square	F-Ratio	Level
FPI	1	25	0,5079221	1,059683	0,48	0,495114
TFP	1	25	0,6426508	1,054294	0,61	0,442286
PCAE	1	25	26,30634	0,02774632	948,10	0,000000
CPF	1	25	0,159483	1,073621	0,15	0,703188
NTF	1	25	0,08730993	1,076508	0,08	0,778152

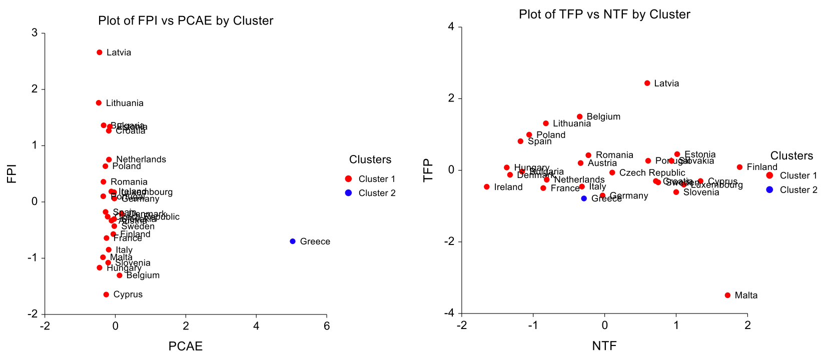
Distance Section for Cluster 1

Row Label	Cluster	Dist1	Dist2
1 Austria	1	0,6476	5,2087
2 Belgium	1	2,5660	5,7149
3 Bulgaria	1	1,7927	5,7503
4 Croatia	1	1,5563	5,5849
5 Cyprus	1	2,1667	5,5337
6 Czech Republic	1	1,0410	5,2668
7 Denmark	1	1,3960	4,9418
8 Estonia	1	1,6830	5,7521
9 Finland	1	2,2537	5,5528
10 France	1	1,2185	5,2605
11 Germany	1	0,8616	5,0974
13 Hungary	1	1,8105	5,5828
14 Ireland	1	2,1508	5,3828
15 Italy	1	1,0348	5,1485
16 Latvia	1	3,6144	7,1880
17 Lithuania	1	2,4172	6,3976
18 Luxembourg	1	1,5603	5,4343
19 Malta	1	4,2565	6,5321
20 Netherlands	1	1,2445	5,4452
21 Poland	1	1,6263	5,7802
22 Portugal	1	0,6749	5,5337
23 Romania	1	2,3728	5,8210
24 Slovakia	1	2,6790	5,6332
25 Slovenia	1	1,6711	5,3882
26 Spain	1	1,6869	5,6899
27 Sweden	1	1,1354	5,2020
Count = 26			

Distance Section for Cluster 2

Row Label	Cluster	Dist1	Dist2
12 Greece	2	5,2634	0,0000
Count = 1			

Fig. 10. K-means cluster analysis report (2 clusters) in NCSS

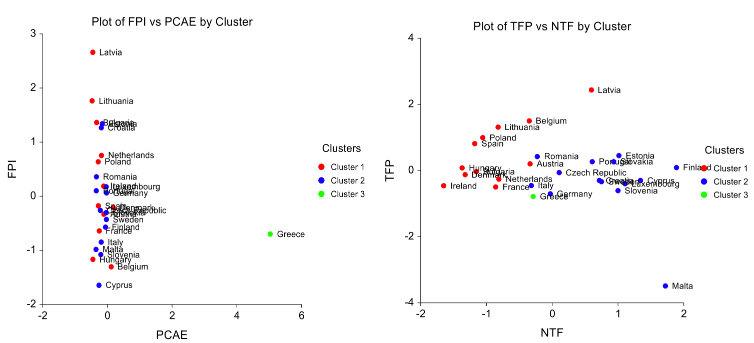


(a) (b)

Fig. 11. K-means clustering (2 clusters) in NCSS
along the axes *FPI*/*PCAE* (a), and *TFP*/*NTF* (b)

Due to k-means method, we understood that the reason for this was an extremely high value of the *PCAE* factor, as can be seen from Fig. 11.

Therefore, we repeated the k-means algorithm for 3 (Figs. 12, 13) and 4 clusters (Figs. 14, 15). Fig. 12 shows the results of clustering along the *FPI* and *PCAE* axes, as well as *TFP* and *NTF*, as in Fig. 11, and Fig. 15 – along the axes *CPF* and *NTF*, and also *PCAE* and *NTF*, respectively.



(a) (b)

Fig. 12. K-means clustering (3 clusters) in NCSS
along the axes *FPI*/*PCAE* (a), and *TFP*/*NTF* (b)

Variables	Cluster1	Cluster2	Cluster3
FPI	0,293665	-0,2017583	-0,6993641
TFP	0,496183	-0,3691091	-0,7866695
PCAE	-0,2175404	-0,1730435	5,033094
CPF	0,3416536	-0,2648539	-0,391888
NTF	-0,8598408	0,7577178	-0,289959
Count	12	14	1

Distance Section for Cluster 1

Row Label	Cluster	Dist1	Dist2	Dist3
1 Austria	1	0,8594	1,3838	5,2087
2 Belgium	1	2,3328	3,0549	5,7149
3 Bulgaria	1	1,4122	2,4552	5,7503
7 Denmark	1	1,0896	2,0906	4,9418
10 France	1	1,3419	1,7271	5,2605
13 Hungary	1	1,6167	2,3699	5,5828
14 Ireland	1	2,0368	2,6084	5,3828
16 Latvia	1	3,3490	4,0518	7,1880
17 Lithuania	1	1,7314	3,1703	6,3976
20 Netherlands	1	0,9027	1,9866	5,4452
21 Poland	1	0,6629	2,5109	5,7802
26 Spain	1	0,8782	2,5283	5,6899
Count = 12				

Distance Section for Cluster 2

Row Label	Cluster	Dist1	Dist2	Dist3
4 Croatia	2	2,1818	1,4788	5,5849
5 Cyprus	2	3,0768	1,5405	5,5337
6 Czech Republic	2	1,7948	1,0103	5,2668
8 Estonia	2	2,1672	1,7330	5,7521
9 Finland	2	3,2151	1,5303	5,5528
11 Germany	2	1,4670	1,1415	5,0974
15 Italy	2	1,5957	1,2551	5,1485
18 Luxembourg	2	2,2458	1,4026	5,4343
19 Malta	2	4,9782	3,7649	6,5321
22 Portugal	2	1,4821	0,8557	5,5337
23 Romania	2	2,6999	2,4447	5,8210
24 Slovakia	2	3,3758	2,3086	5,6332
25 Slovenia	2	2,5230	1,2244	5,3882
27 Sweden	2	1,9493	0,9520	5,2020
Count = 14				

Distance Section for Cluster 3

Row Label	Cluster	Dist1	Dist2	Dist3
12 Greece	3	5,4693	5,2516	0,0000
Count = 1				

Fig. 13. K-means cluster analysis report (3 clusters) in NCSS

Variables	Cluster1	Cluster2	Cluster3	Cluster4
FPI	-0,6993641	-0,7561227	-0,2770992	0,9204657
TFP	-0,7866695	-0,841252	0,09861384	0,5277144
PCAE	5,033094	-0,155691	-0,1430903	-0,2805505
CPF	-0,391888	0,3805785	0,3118291	-0,5913004
NTF	-0,289959	1,300945	-0,8439647	0,1964335
Count	1	6	11	9

Distance Section for Cluster 1

Row Label	Cluster	Dist1	Dist2	Dist3	Dist4
12 Greece	1	0,0000	5,3800	5,2440	5,6254
Count = 1					

Distance Section for Cluster 2

Row Label	Cluster	Dist1	Dist2	Dist3	Dist4
5 Cyprus	2	5,5337	1,2962	2,6629	2,8799
9 Finland	2	5,5528	1,8656	3,0584	2,3323
18 Luxembourg	2	5,4343	1,2317	2,1571	2,2031
19 Malta	2	6,5321	2,8986	4,5618	5,0818
25 Slovenia	2	5,3882	0,5201	2,1061	2,6428
27 Sweden	2	5,2020	0,8574	1,6676	2,0867
Count = 6					

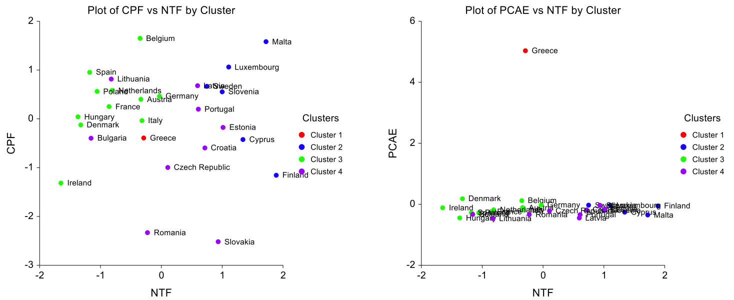
Distance Section for Cluster 3

Row Label	Cluster	Dist1	Dist2	Dist3	Dist4
1 Austria	3	5,2087	1,9513	0,5194	1,6919
2 Belgium	3	5,7149	3,1332	2,2218	3,3109
7 Denmark	3	4,9418	2,7875	0,7477	2,0701
10 France	3	5,2605	2,1553	0,6957	2,2627
11 Germany	3	5,0974	1,5457	1,1850	1,8315
13 Hungary	3	5,5828	2,8356	1,0913	2,6782
14 Ireland	3	5,3828	3,4865	1,9196	2,2967
15 Italy	3	5,1485	1,6866	0,9966	2,1219
20 Netherlands	3	5,4452	2,6173	1,1040	1,7135
21 Poland	3	5,7802	3,2416	1,3026	1,7552
26 Spain	3	5,6899	3,0336	1,0106	2,3143
Count = 11					

Distance Section for Cluster 4

Row Label	Cluster	Dist1	Dist2	Dist3	Dist4
3 Bulgaria	4	5,7503	3,3716	1,7935	1,5143
4 Croatia	4	5,5849	2,3384	2,3633	1,0219
6 Czech Republic	4	5,2668	2,0062	1,5965	1,3606
8 Estonia	4	5,7521	2,4922	2,4874	0,9993
16 Latvia	4	7,1880	4,7142	3,9726	2,8555
17 Lithuania	4	6,3976	3,8979	2,4025	2,0530
22 Portugal	4	5,5337	1,5542	1,4985	1,2165
23 Romania	4	5,8210	3,4777	2,7604	1,8459
24 Slovakia	4	5,6332	3,0990	3,2846	2,3798
Count = 9					

Fig. 14. K-means cluster analysis report (4 clusters) in NCSS



(a) (b)

Fig. 15. K-means clustering (4 clusters) in NCSS along the axes *CPF*/*NTF* (a), and *PCAE*/*NTF* (b)

As a result, we observed a problem with *PCAE* factor for Greece (which again singled out in a separate cluster). So, the formation of a separate cluster from only one country does not indicate a successful solution.

Having confirmed that increased number of clusters for the k-means method does not eliminate the problem of forming a cluster consisting of a single country, we decided to proceed with maximum quantity of 3 clusters based on this method for the ease of interpretation and the need for further analysis by k-medoids (Medoid Partioning), the results of which are shown in Figs. 16, 17.

In Fig. 16 countries are grouped into clusters with their identification number. The columns indicate the following characteristics. Nearest Neighbor is the identification number of the closest neighboring cluster to the corresponding country. Average Distance Within is the average distance between this country and all other countries in the cluster. Average Distance Neighbor is the average distance between this country and the countries in the nearest neighbor. These are the values for computation of the silhouette.

A Silhouette Value (SV) is constructed for each country, $SV \in [-1; 1]$. SV measures how well an object has been classified by comparing its dissimilarity within its cluster to its dissimilarity with its nearest neighbor. If $SV \to 1$, the country is well classified. It means that its dissimilarity with other objects in its cluster is much less than its dissimilarity with objects in the nearest cluster.

If $SV \rightarrow 0$, the country was just between two clusters and the country was arbitrarily assigned to this cluster. If $SV \rightarrow -1$, the country is poorly classified.

Silhouette Bar is a bar chart of the silhouette values sorted by cluster number and silhouette value. The more bars the better structure of the cluster.

Row	Cluster	Nearest Neighbor	Average Distance Within	Average Distance Neighbor	Silhouette Value	Silhouette Bar
1 Austria	1	3	20,98	33,55	0,3746	|IIIIIIIIIIIIIIIIIII
2 Belgium	1	2	38,39	51,60	0,2561	|IIIIIIIIIIIII
3 Bulgaria	1	3	28,89	36,83	0,2156	|IIIIIIIIIII
7 Denmark	1	3	21,97	39,38	0,4420	|IIIIIIIIIIIIIIIIIIIIII
10 France	1	3	22,18	39,69	0,4413	|IIIIIIIIIIIIIIIIIIIIII
11 Germany	1	3	25,42	34,17	0,2560	|IIIIIIIIIIIII
13 Hungary	1	3	26,25	44,95	0,4161	|IIIIIIIIIIIIIIIIIIIII
14 Ireland	1	3	32,52	41,21	0,2110	|IIIIIIIIIII
15 Italy	1	3	24,79	35,61	0,3039	|IIIIIIIIIIIIIII
17 Lithuania	1	3	35,73	42,81	0,1653	|IIIIIIII
20 Netherland	1	3	23,01	37,36	0,3840	|IIIIIIIIIIIIIIIIIII
21 Poland	1	3	24,52	39,12	0,3731	|IIIIIIIIIIIIIIIIIII
26 Spain	1	3	23,81	43,77	0,4560	|IIIIIIIIIIIIIIIIIIIIIII
Cluster Average	1	(13)	26,80	40,00	0,3304	
5 Cyprus	2	3	40,50	38,79	-0,0422	|
12 Greece	2	1	78,16	76,41	-0,0224	|
18 Luxembourg	2	1	36,90	34,89	-0,0545	|
19 Malta	2	1	56,10	67,39	0,1676	|IIIIIIII
25 Slovenia	2	1	32,81	35,68	0,0804	|IIII
27 Sweden	2	1	33,10	29,89	-0,0969	|
Cluster Average	2	(6)	46,26	47,18	0,0053	
8 Estonia	3	1	28,97	37,09	0,2188	|IIIIIIIIIII
4 Croatia	3	1	28,54	35,62	0,1988	|IIIIIIIIII
6 Czech Repu	3	1	28,47	28,72	0,0086	|
9 Finland	3	2	33,92	42,09	0,1941	|IIIIIIIIII
16 Latvia	3	1	53,45	54,37	0,0169	|I
22 Portugal	3	1	29,58	27,29	-0,0773	|
23 Romania	3	1	34,40	41,41	0,1694	|IIIIIIII
24 Slovakia	3	1	35,04	49,30	0,2893	|IIIIIIIIIIIIII
Cluster Average	3	(8)	34,05	39,49	0,1273	
Overall Average		(27)	33,27	41,44	0,1980	= SC
Maximum Distance		3,275787				

Fig. 16. K-medoids cluster analysis report (3 clusters) in NCSS

This method introduces a silhouette value is for each object, ranging from 1 to -1. Silhouette value measures how well an object has been classified by comparing its dissimilarity within its cluster to its dissimilarity with its nearest neighbor. The majority of country values are positive (signifying that the objects are well classified). Also, there are some negative values that indicated a possibility of further tossing out of cluster configuration. But we decided not to do this, as far as we understand the limitations of our input data taken only for 2016.

Fig. 17 provides the overall information concerning iterative process of k-medoids clustering and parameters of every indecies of each obtained cluster.

Variables		FPI, TFP, PCAE, CPF, NTF	
Method: Spath,		Objective Function: Silhouette	
Distance Type: Euclidean,		Scale Type: None	
Iteration Detail Section			
	(Minimize This)	Adjusted	(Maximize This)
Number	Average	Average	Average
Clusters	Distance	Distance	Silhouette
3	199,228061	22,136451	0,106349
3	120,902128	13,433570	0,159642
3	123,432443	13,714716	0,157911
3	118,648456	13,183162	0,121021
3	122,447703	13,605300	0,197984
3	122,175417	13,575046	0,169657
Iteration Summary Section			
	(Minimize This)	Adjusted	(Maximize This)
Number	Average	Average	Average
Clusters	Distance	Distance	Silhouette
3	122,447703	13,605300	0,197984
Cluster Medoids Section			
Variable	Cluster1	Cluster2	Cluster3
FPI	-0,3321139	-1,079073	-0,2615117
TFP	0,2024935	-0,608986	-0,06294302
PCAE	-0,10794	-0,2029022	-0,2173879
CPF	0,4005181	0,5536308	-0,9968126
NTF	-0,336809	1,00009	0,1059177
Row	1 Austria	25 Slovenia	6 Czech Repu

Fig. 17. The overall information about the iterative process of k-medoids clustering and the parameters of 3 obtained clusters

Based on the information obtained, we identified 3 clusters (Fig. 18). We consider this solution successful, because with increasing number of clusters, the algorithm also began to give options where the cluster was formed by one country. Thus, k-medoids method turned out to be the most adequate in our research compared to other methods.

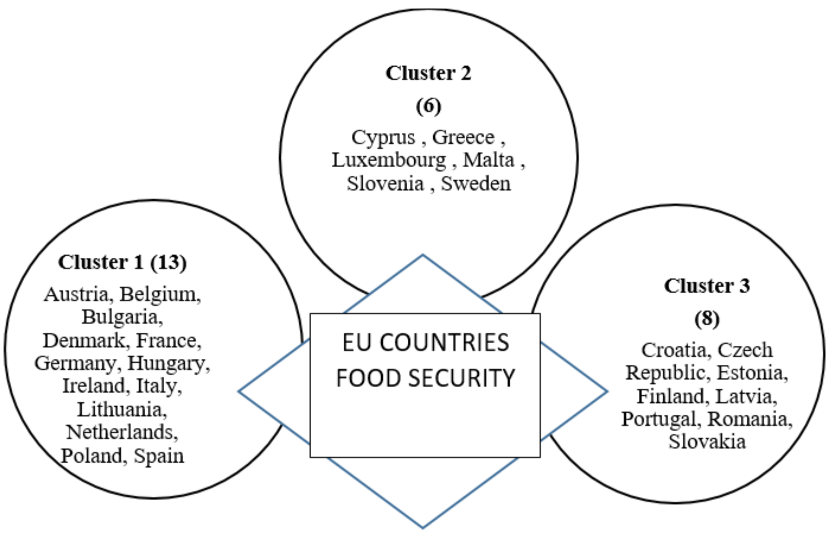


Fig. 18. EU countries clusterization for the food security state

Cluster 1 included the countries (Fig. 19):

1) mainly with a high level of FPI index (for 9 countries > 100);

2) with a high level of TFP growth (for all countries);

3) with a significant reduction in per capita agriculture expenditures (for all countries except Denmark and Belgium);

4) with an annual increase in consumer food prices of up to 2.24% (except for Bulgaria and Ireland, where food prices have fallen);

5) mostly net food importers (except Germany, Italy, Austria and Belgium, which are net food exporters).

Cluster 2 included the countries (Fig. 20):

1) mainly with an insufficient level of the FPI index (except for Luxembourg);

	Country	FPI		Country	TFP		Country	PCAE
17	Lithuania	136,60	2	Belgium	131,75	7	Denmark	122,45
3	Bulgaria	129,85	17	Lithuania	129,16	2	Belgium	113,07
20	Netherlands	119,58	21	Poland	124,78	11	Germany	90,86
21	Poland	117,60	26	Spain	122,28	1	Austria	78,43
14	Ireland	110,02	1	Austria	113,88	14	Ireland	77,49
11	Germany	107,92	13	Hungary	112,16	20	Netherlands	67,27
26	Spain	103,90	3	Bulgaria	110,59	15	Italy	66,14
7	Denmark	103,33	7	Denmark	109,36	10	France	56,78
1	Austria	101,29	20	Netherlands	107,48	26	Spain	52,73
10	France	96,07	15	Italy	104,82	21	Poland	51,98
15	Italy	92,59	14	Ireland	104,76	3	Bulgaria	44,42
13	Hungary	87,20	10	France	104,24	13	Hungary	26,60
2	Belgium	84,89	11	Germany	101,39	17	Lithuania	23,06

	Country	CPF		Country	NTF
2	Belgium	102,24	11	Germany	0,085
26	Spain	101,40	15	Italy	0,022
17	Lithuania	101,23	1	Austria	0,018
20	Netherlands	100,95	2	Belgium	0,015
21	Poland	100,93	20	Netherlands	-0,086
11	Germany	100,81	17	Lithuania	-0,089
1	Austria	100,73	10	France	-0,097
10	France	100,55	21	Poland	-0,140
13	Hungary	100,30	3	Bulgaria	-0,161
15	Italy	100,21	26	Spain	-0,166
7	Denmark	100,10	7	Denmark	-0,198
3	Bulgaria	99,77	13	Hungary	-0,209
14	Ireland	98,67	14	Ireland	-0,270

Fig. 19. The parameters of factors for cluster 1

	Country	FPI		Country	TFP		Country	PCAE
18	Luxembourg	109,81	5	Cyprus	106,95	12	Greece	861,99
27	Sweden	99,64	27	Sweden	106,44	27	Sweden	90,40
12	Greece	95,10	18	Luxembourg	105,58	18	Luxembourg	89,47
19	Malta	90,31	25	Slovenia	102,70	25	Slovenia	63,96
25	Slovenia	88,70	12	Greece	100,25	5	Cyprus	55,43
5	Cyprus	79,14	19	Malta	63,00	19	Malta	41,45

	Country	CPF		Country	NTF
19	Malta	102,16	19	Malta	0,469
18	Luxembourg	101,53	5	Cyprus	0,386
27	Sweden	101,05	18	Luxembourg	0,335
25	Slovenia	100,92	25	Slovenia	0,311
12	Greece	99,78	27	Sweden	0,256
5	Cyprus	99,74	12	Greece	0,028

Fig. 20. The parameters of factors for cluster 2

2) with a high level of TFP growth (except for Malta, which lags significantly behind others);

3) with a characteristic decrease in per capita agricultural expenditures (except for Greece, for which this indicator has increased significantly);

4) with an annual increase in consumer food prices of up to 2.16% (except for Greece and Cyprus, where food prices have fallen);

5) net food exporters.

Cluster 3 included countries (Fig. 21):

1) mainly with a high level of FPI index (for 5 out of 6 countries more than 100);

2) with a high level of TFP growth (for all countries);

3) with a significant reduction in per capita agriculture expenditures (for all countries);

4) with an annual reduction in consumer food prices of up to 3% (except for Latvia and Portugal, where food prices have risen);

5) net food exporters.

	Country	FPI
16	Latvia	151,72
4	Croatia	128,23
23	Romania	112,93
22	Portugal	108,59
6	Czech Republic	102,48
24	Slovakia	101,77
9	Finland	97,26

	Country	TFP
16	Latvia	144,64
23	Romania	116,88
22	Portugal	114,75
24	Slovakia	114,73
9	Finland	112,31
6	Czech Republic	110,22
4	Croatia	106,95

	Country	PCAE
24	Slovakia	89,67
9	Finland	86,21
4	Croatia	66,55
6	Czech Republic	61,75
23	Romania	43,44
22	Portugal	42,69
16	Latvia	26,43

	Country	CPF
16	Latvia	101,07
22	Portugal	100,49
4	Croatia	99,53
6	Czech Republic	99,05
9	Finland	98,86
23	Romania	97,44
24	Slovakia	97,22

	Country	NTF
9	Finland	0,506
24	Slovakia	0,297
4	Croatia	0,249
22	Portugal	0,225
16	Latvia	0,223
6	Czech Republic	0,115
23	Romania	0,042

Fig. 21. The parameters of factors for cluster 3

We note that the EU as a whole is characterized by a reduction in support for the agricultural sector (except for Greece). Interestingly, the vast majority of small countries in which agriculture is not a priority sector are, in fact, net food exporters. Whereas such countries with developed agro-industrial complex as France, the Netherlands, Spain, Poland – are, in fact, net food importers. We also observe the relative food price stability for consumers (inflation within 2.5%), and in some EU countries (Ireland, Greece, Slovakia) even food prices reduction. Despite the declining trend in overall factor productivity in the EU, the growth rate of TFP demonstrates positive dynamics, except for Malta. But for 10 EU countries, the FPI indicator shows an insufficient level of food production and the need to import food. Therefore, given the above, we can assume that EU food security level is sufficient. However, according to the results of cluster analysis, we observe a certain heterogeneity in the state of food security in 3 specific groups of EU countries.

Our policy recommendation are summarized in Table 3.

Table 3

POLICY RECOMMENDATIONS FOR GROUPS OF EU COUNTIES

Cluster	Policy recommendations	Comments
Cluster 1	Cluster 1 countries should increase their level of food security via increased government support to agriculture with a view to downsizing imports	State of food security is optimal in Germany, Italy, Austria and Belgium
Cluster 2	State of food security is optimal in all Cluster 2 countries	Level of agriculture support in Greece is abnormal – almost 10 times higher than in Luxembourg
Cluster 3	Cluster 3 countries should follow their current agripolicy aimed at diminishing state support	All these countries have the highest level of food security in the EU

The obtained results gave us an opportunity to define the advantages and disadvantages of hierarchical and non-hierarchical clustering methods applied in our research (Table 4).

Thus, we consider hierarchical SOM and dendrograms methods perfect for data visualization, whereas by using k-means and k-medoids methods we presented more accurate and detailed solutions, as presented in Table 4.

Table 4

THE ADVANTAGES AND DISADVANTAGES OF HIERARCHICAL AND NON-HIERARCHICAL CLUSTERING METHODS APPLIED AS PERTINENT TO OUR RESEARCH

Methods		Advantages	Disadvantages
hierarchical	dendrograms	• perfect data visualization	• it is necessary to set the number of clusters • the presence of outliers
	SOM	• perfect data visualization • use of universal approximator • self-organization of the network	• it is necessary to set the number of clusters • prolonged process • the presence of outliers
non-hierarchical	k-means	• ease of use • speed of use • clarity and transparency of the algorithm	• it is necessary to set the number of clusters • the presence of outliers
	k-medoids	• ease of use • speed of use • clarity and transparency of the algorithm • k-medoids is less sensitive to outliers than k-means = more accurate results	• it is necessary to set the number of clusters

Conclusions

In the paper we considered the theoretical and methodological tools of cluster analysis. Particular attention was given to justifying the choice of factors to study the state of EU food security. Finally, as a result of applying 4 methods of cluster analysis – SOM, dendrograms, k-means and k-medoids, we classified the EU countries into 3 groups for the state of food security.

The obtained results gave us an opportunity to define the advantages and disadvantages of the selected clustering methods. Thus, we consider SOM and dendrograms methods perfect for data visualization, whereas k-means and k-medoids give more adequate and detailed solutions.

Overall, the state of food security at EU level is optimal, however as the cluster analysis has proved there is some heterogeneity between countries, which is significantly reduced within separate clusters.

So, one cluster of EU countries mainly consists of net food importers. This group is characterized by a high level of FPI index and TFP growth, reduced agricultural support, and annual increase in consumer food prices. Therefore, the countries of the first cluster should increase their level of food security via rising government support for agriculture with a view to downsizing imports.

The second group of countries includes net food exporters with insufficient FPI, high TFP growth, reducing agricultural support, and increasing consumer food prices. State of food security is optimal in all countries of this cluster.

The third group includes net food exporters with high FPI, high TFP growth, a characteristic decrease in agricultural support and diminishing consumer food prices. Thus, the countries of the third cluster should follow their current agripolicy aimed at reducing state support.

As follows from the study, cluster analysis is a universal and powerful tool for economic and mathematical modeling, which allowed revealing the groups of countries with different levels of food security in order to identify adequate strategies for ensuring food security, depending on the initial conditions of these countries.

References

1. Sen, A. (1983). *Poverty and famines: An essay on entitlement and deprivation*. Oxford university press. https://doi.org/10.1093/0198284632.001.0001

2. Friedmann, H., & McMichael, P. (1989). Agriculture and the state system: the rise and fall of national agricultures, 1870 to the present. *Sociologia Ruralis, 29*(2), 93-117. https://doi.org/10.1111/j.1467-9523.1989.tb00360.x

3. FAO. (1996, November 13-17). *Rome Declaration on World Food Security*. https://www.fao.org/3/w3613e/w3613e00.htm

4. Committee on World Food Security. (2012). *CFS 39 Final Report* (CFS 2012/39/4). FAO. http://www.fao.org/3/MD776E/MD776E.pdf

5. IFC. (2021). *International Food Safety Forum*. https://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/agribusiness/events/food+safety+forum

6. Buiak, L., Bashutska, O., Pryshliak, K., Hryhorkiv, V., Hryhorkiv, M., & Kobets, V. (2020). Models of Rental Payments Formation for Agricultural Land Plots Taking into Account the Ecological Level of Economy. In *Proceedings of 2020 10th International Conference on Advanced Computer Information Technologies* (pp. 204-208). IEEE. https://doi.org/10.1109/ACIT49673.2020.9208959

7. Wittman, H. (2011). Food Sovereignty: A new Rights Framework for Food and Nature? *Environment and Society*, *2*(1), 87-105. https://doi.org/10.3167/ares.2011.020106

8. Maxwell, S. (2001). The Evolution of Thinking about Food Security. In S. Devereux, & S. Maxwell (Eds.), *Food Security in Sub-Saharan Africa* (pp. 13-31). ITDG Publishing. https://doi.org/10.3362/9781780440170.002

9. Novikov, Ye.Yu. (2011). Zabezpechennia hlobalnoi prodovolchoi bezpeky v umovakh pryskorennia klimatychnykh zmin [Ensuring global food security in the context of accelerating climate change]. *Kultura narodov Prychernomoria (Culture of the peoples of the Black Sea region)*, *216*, 78-81. http://dspace.nbuv.gov.ua/handle/123456789/64963 [in Ukrainian]

10. Fischler, F., Wilkinson, D., & Benton, T., et al. (2015). *The role of research in global food and nutrition security*. Expo 2015 EU Scientific Steering Committee. https://europa.eu/expo2015/sites/default/files/files/FINAL_Expo-Discussion-paper_lowQ(1).pdf

11. Eurostat. (2017, October 17). *EU trade in food*. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/edn-20171016-1

12. Kaminskyi, A., Nehrey, M., & Komar, M. (2019). Agro-economic models: a review and directions for research. *Periodicals of Engineering and Natural Sciences*, *7*(2), 702-711. https://doi.org/10.21533/pen.v7i2.579

13. Kaminskyi, A. B., Nehrey, M. V., & Zomchak, L. M. (2021). COVID-19: crisis or new opportunities time for the agricultural sector of Ukraine. *IOP Conference Series: Earth and Environmental Science*, *628*(1), Article 012031. https://doi.org/10.1088/1755-1315/628/1/012031

14. Hossaina, M., Mullally, C., & Asadullah, M.N. (2019). Alternatives to calorie-based indicators of food security: An application of machine learning methods. *Food Policy*, *84*, 77–91. https://doi.org/10.1016/j.foodpol.2019.03.001

15. Nica-Avram, G., Harvey, J., Smith, G., Smith, A., & Goulding, J. (2021). Identifying food insecurity in food sharing networks via machine learning. *Journal of Business Research*, *131*, 469–484. https://doi.org/10.1016/j.jbusres.2020.09.028

16.Kozlovskyi, S., Khadzhynov, I., Mohylova, A., Tomchuk, V., & Kozlovskyi, A. (2018). The concept model for ensuring the economic security in the agrarian sector. *Management theory and Studies for Rural Business and Infrastructure Development*, *40*(4), 531-547. https://doi.org/10.15544/mts.2018.47

17.Tan, P., Steinbach, M., & Kumar, V. (2005). Cluster analysis: basic concepts and algorithms. In *Introduction to data mining* (1st ed., pp. 487-568). Pearson Addison Wesley. https://www-users.cse.umn.edu/~kumar001/dmbook/ch8.pdf

18.Pistunov, I.M., Antoniuk., O., & Turchaninova, I. (2008). *Klasternyi analiz v ekonomitsi [Cluster analysis in economics]*. National Mining University. [in Ukrainian]

19.Erickson, K. (2018). *Introduction to Cluster Analysis*. https://cnls.lanl.gov/external/qbio2018/Slides/Cluster_Lab_June18/qBio_Clustering_Lecture.pdf

20.Kobets, V., Yatsenko, V., & Voynarenko, M. (2019). Cluster analysis of countries inequality due to IT development. *CEUR Workshop Proceedings*, *2393*, 406–421. http://ceur-ws.org/Vol-2393/paper_341.pdf

21.Kobets, V., Yatsenko, V., & Voynarenko, M. (2020). Cluster Analysis of Countries Inequality Due to IT Development Through Macros Application. In V. Ermolayev, F. Mallet, V. Yakovyna, H. Mayr, & A. Spivakovsky (Eds.), *Communications in Computer and Information Science: Vol. 1175. Information and Communication Technologies in Education, Research, and Industrial Applications* (pp. 415–439). Springer. https://doi.org/10.1007/978-3-030-39459-2_19

22.Tishchenko, O.M. (2010). Klastery yak vektor rozvytku ekonomiky: orhanizatsiia, sutnist i kontseptsii [Clusters as a vector of economic development: organization, essence and concepts]. *Teoretychni ta prykladni pytannia ekonomiky (Theoretical and applied issues of economics)*, *21*, 74-80. http://tppe.econom.univ.kiev.ua/data/2010_21/Zb21_10.pdf [in Ukrainian]

23.Jipkate, B.R., & Gohokar, V.V. (2012). A comparative analysis of fuzzy c-means clustering and k means clustering algorithms. *International Journal of Computational Engineering Research*, *2*(3), 737-739. http://www.ijceronline.com/papers/Vol2_issue3/T023737739.pdf

24.Hartigan, J.A., & Wong, M.A. (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

25.NCSS. (2020). K-Means Clustering. Chapter 446. In *NCSS Statistical Software* (pp. 446-1–446-8). NCSS. Retrieved July 1, 2020, from https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/K-Means_Clustering.pdf

26.NCSS. (2020). Medoid Partitioning. Chapter 447. In *NCSS Statistical Software* (pp. 447-1–447-11). NCSS. Retrieved July 1, 2020, from https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Medoid_Partitioning.pdf

27.FAOSTAT. (2020). *Suite of Food Security Indicators* [Data set]. Retrieved July 1, 2020, from http://www.fao.org/faostat/en/#data/FS

28.Van Dijk, M., & Meijerink, G. (2014). A review of global food security scenario and assessment studies: results, gaps and research priorities. *Foodsecure*, *February*, Article 20. https://www.wecr.wur.nl/WECRGeneral/FoodSecurePublications/20_Dijk_Meijerink_review-global-food.pdf

29.World Bank. (2020). *Food production index* [Data set]. Retrieved July 1, 2020, from https://data.worldbank.org/indicator/AG.PRD.FOOD.XD

30.European Commission. (2020). *CAP CONTEXT INDICATORS – 2019 update* [Data set]. Retrieved July 1, 2020, from https://agriculture.ec.europa.eu/document/download/fd20e275-115a-4bdd-af9e-364650794270_en?filename=cap-context-indicators-table_2019_en_0_0.pdf

31.Van Ark, B., de Vries, K., & Jäger, K. (2018). Is Europe's Productivity Glass Half Full or Half Empty? *Intereconomics*, *53*(2), 53-58. https://doi.org/10.1007/s10272-018-0721-z

32.International Food Policy Research Institute. (2018). Statistics on Public Expenditures for Economic Development (SPEED). In *Food policy indicators: tracking change* (pp. 106-111). USAID. http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/132327/filename/133362.pdf

33.Google Drive. (2021, August 1). *Data set for Cluster analysis for the State of Food Security* [Data set]. https://drive.google.com/file/d/1dVpjcBDYpDi05r77pw5bb7CmLBEgs8wG/view?usp=sharing

The article was submitted on 2021, August 8