УДК 339.7:622
DOI: 10.32342/2074-5354-2022-2-57-3

L. HARMIDER,
Dr. Sc. (Econ.), prof., SHEI Ukrainian State Chemical Technology University, Dnipro (Ukraine)
https://orcid.org/0000-0001-7837-2734

S. FEDULOVA,
Dr. Sc. (Econ.), prof., Alfred Nobel University, Dnipro (Ukraine)
https://orcid.org/0000-0002-5163-3890

YU. BARTASHEVSKA,
PhD in Economics, associate professor, Alfred Nobel University, Dnipro (Ukraine)
https://orcid.org/0000-0002-0300-0693

V. KOMIRNA,
Dr. Sc. (Econ.), prof., European University Servant of God Robert Schuman, Radom (Poland)
https://orcid.org/0000-0002-9298-3010

ASSESSING THE REGIONAL LABOR MARKET BY USING DATA MINING METHODS: WAYS OF EFFECTIVE FUNCTIONING

As a result of the uneven development of certain territories, it is more feasible and effective to tackle the practical issues of labor market regulation at the regional level. This ensures sufficient regulation of the system. Since it is necessary to properly account for the regional differences in practice, it is required that these issues be methodologically justified. Therefore, the aim of this paper is to investigate regional labor markets based on indicators of the socio-economic development of regions using the data mining methods.

The current study has clustered regions of Ukraine on the basis of the level of their socio-economic development using data mining methods, in particular Kohonen maps and the k-means methods. Based on the research results, it has been proven that gross regional product, wages, and investments are crucial indicators for clustering the regions in terms of labor market development. The authors state that the assessment of the regional labor market in terms of socio-economic development testifies to the prevalence of extensive factors in the development of most regions in Ukraine. Based on the broad comprehensive typology of regional labor markets constructed by the authors, they propose to identify the priority areas for their regulation, whose application is most effective for a specific group of labor markets under investigation. The authors propose to consider the following areas as priorities: attracting investments by legal entities and individuals; arranging for special types of work (public works); supporting self-employment and small businesses, including the legalization of the unregistered shadow employment of economically active population; enhancing and expanding professional training and retraining of unemployed individuals to assist them in their search for for new careers and occupations; creating an effective system of labor market monitoring. The study of the regional labor markets by using intellectual data mining methods has allowed for generalization of the main proposals and recommendations related to the regulation of regional labor markets.

Key words: labor market, region, data mining methods, indicators, socio-economic development.

Внаслідок нерівномірності розвитку окремих територій практичні питання регулювання ринку праці реальніше й ефективніше вирішувати на регіональному рівні. Це забезпечує достатню врегульованість системи. Оскільки на практиці треба грамотно враховувати регіональні відмінності, то необхідно і методологічно розробляти ці питання. Тому, метою статті є дослідження регіональних ринків праці за показниками соціально-економічного розвитку регіонів з використанням методу інтелектуального аналізу бази даних. У дослідженні здійснено кластеризацію регіонів країни за рівнем соціально-економічного розвитку методами Data Mining, за допомогою карт Кохонена та методів k-means. За результатами дослідження, доведено, що показники валового регіонального продукту, заробітної плати та інвестицій є визначальними для кластеризації областей України за рівнем розвитку ринку праці. На думку авторів, оцінка регіонального ринку праці за рівнем соціально-економічного розвитку свідчить про перевагу екстенсивних факторів у розвитку більшості регіонів України. За результатами створення широкої комплексної типології регіональних ринків праці, автори пропонують виявити пріоритетні напрямки їхнього регулювання, використання яких найбільш ефективно для конкретної групи досліджуваних ринків праці. Запропоновано, першочерговими вважати заходи з залучення інвестицій юридичних та фізичних осіб; організації спеціальних видів робіт (суспільні роботи), розвитку самозайнятості і малого бізнесу, зокрема легалізація незареєстрованої тіньової зайнятості економічно активного населення; активізація і розширення професійної підготовки і перепідготовки незайнятого населення на нові професії; створення дієвої системи моніторингу ринку праці. Дослідження регіональних ринків праці з використанням методів інтелектуального аналізу даних дає нам можливість узагальнити основні пропозиції і рекомендації з регулювання регіональних ринків праці.

Ключові слова: ринок праці, регіон, метод Data Mining, показники, соціально-економічний розвиток.

Introduction and problem statement. Given the uneven development of certain territories, it is more feasible and effective to resolve the practical issues of labor market regulation at the regional level. This ensures sufficient regulation of the system. Since it is necessary to properly account for the regional differences in practice, it is required that these issues be methodologically justified. A defined step in this direction was to develop a methodology to quantify the level of development of the regional labor market based on the integrated indicators for measuring socio-economic development, which allowed to specify trends dominating in the development of the regions – extensive or intensive, which in turn allowed to determine the priority areas for the implementation of state employment policies within a particular group of regions.

The set problem suggests solving the following tasks: to determine the information base, evaluation criteria, and a list of indicators measuring regional labor markets performance; to analyze the state of Ukraine's labor market based on the set of regional indicators and integrated coefficients measuring the levels of extensive and intensive development; to devise a methodology for assessing regional differentiation based on the data mining method; to segment the regions of Ukraine according to the level of socio-economic development by using Kohonen maps; to identify patterns and problems related to Ukraine's regional labor markets.

The object of this study is the economic relations that develop in the process of functioning of the regional labor market. The subject of the study is the regional labor market in Ukraine.

Literature review. An interesting fact is that the resulting benefit from the increased efforts, in terms of higher anticipated income after training, turns out to be less for students with enhanced abilities than for students with low abilities. Analysts also note that broad prospects in the labor market might have a negative impact on students' efforts [1].

As Schwab notes [2], we are at the beginning of the fourth Industrial Revolution, which began at the turn of this century and is based on the digital revolution of the third Industrial Revolution. The fourth industrial revolution can be characterized by much more universal and mobile Internet, smaller and more powerful sensors that became cheaper, and by machine learning based on artificial intelligence. In general, digital technologies have several destructive directions, offering new radical ways of doing business with an emphasis on the lowest skills [3]. Given these driving factors and the changes we face, there is one certainty: new technologies will radically change the nature of work in all industries and occupations [4].

A synthetic control analysis reveals that the influx of repatriates from Mozambique and Angola to Portugal in the mid-1970's exerted a significant adverse effect on the labor market performance. The results show that the Portuguese labor market reacted in the following way: an increase in the number of employees led to a decrease in average productivity and wages [5].

Scientists Ciani, David, Blasio [6] note that mobility of the workforce is now critical to regional convergence at the employment level. There is also evidence indicating that additional resources allocated devoted for the active labor market policy are directly related to the production growth rate. This differential becomes larger during economic upturns when market conditions are improved compared to the trend [7]. In addition, there are facts that the availability of temporary contract jobs increases and their share in the labor market becomes increasingly important as workplace safety improves [8].

As noted by Dengler & Matthes [9], the digital transformation would have a major impact on the labor market. The findings show that half of all current jobs are subject to automation over the next 10–20 years. The scientists also assume that only certain tasks in a profession can be replaced, not the entire professions.

Technological advances are currently transforming the labor market, which leads, on the one hand, to the automation of certain jobs and tasks, and to the emergence of new types of work, on the other hand. Proactive preparation for this new reality requires a deep, detailed understanding of these changes and their impact on jobs and employment. Skills are a new currency in the labor market. Based on these changes, LinkedIn developed Genome Skills, a new metric that makes it possible to apply these analytical possibilities for a deeper understanding of trends and developments in the labor market. The skills that enjoyed the greatest growth can be divided into four categories: functional skills such as marketing and customer service; soft power skills such as leadership; digital skills such as social networks; and additional skills such as the English language [10].

The UK Government Department for Digital, Culture, Media and Sport commissioned to conduct quantitative and qualitative research to better understand the current state of cybersecurity skills in the labor market. This work is part of the UK government's National Cyber Security Strategy in 2016–2021. The £1.9 billion investment under this strategy aims to secure a steady supply of UK cybersecurity professionals to meet the growing demands of an increasingly digital economy in the public, private and defense sectors. Currently, there is no agreed definition of cybersecurity skills or a cybersecurity [11].

When analyzing the problems of the Georgian economy, the Ministry of Economy and Sustainable Development of Georgia also comes to the conclusion that it is necessary to promote the process of mastering entrepreneurial skills by young people in the school process and to promote the development of natural sciences (mathematics, physics, etc.) and technical sciences (engineering, architecture, etc.) engaging young people to gain knowledge in this area in order to develop high-tech products in the country. The development of an entrepreneurial culture among young people, starting from the school bench, could positively influence the creation of successful business organizations [12].

Professional skills are increasingly valued in the labor market. Along with academic qualification, the demand for professionally-focused skills has been steadily increasing. In Europe, a lot of professional programs are completed by obtaining a bachelor's degree. No previous research has investigated potential benefits for a labor market after professional magistracy. The most recent data for the European countries suggest that college diplomas or bachelor diplomas allow for higher earnings, especially by women [13].

Interestingly, the liberalization of trade is likely to have a positive impact on employment in countries with flexible labor markets and vice versaю. Mainly, this is a result of the process of adaptation of the labor market to the openness of trade. Researchers Irène Selwaness & Chahir Zaki [14] have come to the conclusion that the rigidity of the labor market reduces the positive impact of exports on employment. That is, rigid labor markets can limit the ease of creating new jobs.

In this case, as noted by Prescott J.J. & Pyle B. [15], the stronger labor markets seem to have a cause-effect link to the low rate of crimes against property. The scientists reveal the statistically significant positive impact of unemployment on the crime level in real estate and a weaker and less coherent relationship between unemployment and the levels of violent crimes.

It should be noted that integration of the Chinese workforce into the global labor markets has been playing a key role in the global reduction of the share of the labor force since the late 1970s, mainly through the channel of international trade [16].

Recent studies indicate that labor markets in South-Eastern Europe should adapt to rapid changes in the European labor market, where opportunities for lifelong work would be very rare, while the mobility of jobs could be further intensified in the future. To perform a comparative analysis of labor markets in the region of South-Eastern Europe, the cited studies applied two multifactorial methods: factor analysis and analysis of main components. In the authors' opinion, the factor analysis based on an analysis of core components could help reduce a significant number of variables to a meaningful, interpreted, and manageable set of factors [17].

The specific needs of regions, territories, sectors, and enterprises are formed in the labor market. The analysis of the regional labor market suggests the analysis of its levels, structure, employment status, and determining those indicators that constitute the methodological basis of the predictive set of labor market skills. In this case, a factor analysis as part of forecasting current employment by industry could be used to forecast the structure of employment by profession and occupation [18].

In general, the economic literature outlines different ways and methods of assessing the level of socio-economic development of a region by means of mathematical methods. All these methods, however, can be reduced to two groups:

1. Calculating the generalized index of a region's development, on the basis of which the ratings of the regions are built.
2. Using econometric models and their characteristics to quantify the level of development.

Both methods and their modifications have their advantages and disadvantages. The most common approach to assessing the level of regional development is the approach focused on some generalized assessment of the region's development in the form of a certain index, and its comparison with the same generalized estimates of the development of other regions.

The most common groups of index approach methods include [19]:

1. Comprehensive assessment of social and economic development of regions.
2. Quality of life as a complex indicator.
3. The Human Development Index (HDI).
4. The global IMD Competitiveness Index.
5. The Growth Competitiveness Index.
6. Evaluation of the effectiveness of regions' development.
7. Ranking the level of socio-economic development.

However, a variety of index approaches regarding assessing the level of socio-economic development of regions does not make it possible to solve the main task – to ensure objectivity of assessment of this level and to conduct a comparative analysis of uneven regional development.

Researchers attempt to perform such an analysis by using econometric methods of assessing the level of socio-economic development of regions. There are two approaches:

1. The first approach involves attempts to identify causal relationships in regional development underpinned by mathematical description of the identified relationships on the basis of the use of econometric methods.

2. The second approach is to find some structural or other patterns inherent in the regions' development and to model these features without revealing casual relationships between the factors and indicators of the regional economy.

It should be noted that there is a huge number of studies in the Ukrainian economic science that are devoted to the designing the system of econometric models of the development of the regional economy.

The main drawback of these numerous models is the impossibility to incorporate all those factors that affect the results of regional development. In addition, it often happens that a researcher is simply unaware of some factors affecting regional development, and these factors are not integrated into the model. As such, such constructs often do not lead to the expected improvement of modelling quality and analysis of the regional economy.

It is entirely possible that many of the factors incorporated in the models have been shaped by the factors hidden from the researcher. If the researcher reveals these factors, then subsequently a system of econometric models might be designed.

The aim of the paper. The main research goal is to study regional labor markets based on the indicators of the socio-economic development of regions using data mining methods.

The main material of the study. One of the most critical stages in the assessment of Ukraine's regions in terms of socio-economic development by using data mining methods is to determine the information base, criteria of evaluation, and a list of estimates. The main aspects of the socio-economic and demographic development of the regions are characterized by a set of statistical indicators related to four blocks of the key factors:

1. Assessment of the demographic situation in a region.

2. Assessment of the social situation in a region.

3. Assessment of the economic situation in a region.

4. Assessment of the organizational environment in a region.

In this case, each of the defined main blocks is characterized by a different number of indicators (Fig. 1). The number of indicators can be expanded if there is a possibility to use a wider information base. Typically, the application of more indicators, specifically the dynamics of economic and social processes, would make it possible to conduct the required calculations with greater accuracy.

The study, by no means, claims to detect all the dependences in the labor market related to all the above-mentioned factors. Based on public data, given in the statistical yearbook "Ukraine in Figures" (2020), by using mathematical methods (correlation-regression and cluster analysis), we obtained two groups of factors that characterize different aspects of the socio-economic and demographic development (Fig.1).

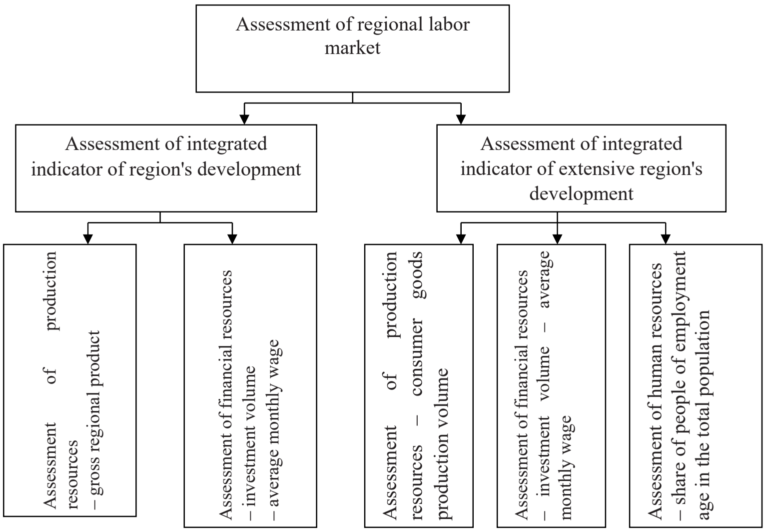
It should be noted that these indicators did not have a statistically significant relationship, so the inclusion any of them in the series of criterion is fully justified. In addition, such a grouping is not a dogma; it can constantly evolve, be supplemented by new indicators that are the most relevant in a given period.

Using these coefficients provides an opportunity to formalize the labor market analysis.

1. The integrated coefficient of the intensive development of regions (I_{int}) is proposed to be determined as the product of indexes of gross regional product (I_{GRP}), investments (I_i), and wages (I_w):

$$I_{int} = I_{GRP} * I_i * I_w , \qquad (1)$$

where $I_{GRP,}$ $I_{i,}$ I_w are the indexes of gross regional product, investments, and wages, calculated relative to the corresponding maximum value.



Source: author's proposition

Fig. 1. General assessment of a regional labor market

2. The integrated coefficient of the extensive development of regions (Iext) is proposed to be determined as the product of the indexes of gross regional product (I_{GRP}), wages (I_w), and the share of people of working age (I_q):

$$I_{ext} = I_{GRP} * I_w * I_q \, , \qquad (2)$$

where I_q is the proportion of people of working age in the total population.

Thus, in addition to the integrated coefficient of the regions' development, which combines the influence of all factors, we calculated individual indices for each indicator based on their absolute values in 2020 using a specific index form, which makes it possible to calculate the integrated development indices for each region (refer to Table 1). It should be noted that I_{GRP}, $I_{i,}$ I_w, I_q were calculated as the ratio of the absolute value of the corresponding indicator to its maximum value by regions.

Table 1

Indices of regions' indicators and the integrated coefficients of the levels of extensive and intensive development in 2020[1]

REGIONS	I_{GRP}	Ii	I_w	I_q	Iint	Iext
Vinnyts′ka	0,278203	0,250553	0,681928	0,957932	0,047534	0,181733
Volyns′ka	0,147245	0,192728	0,660843	0,959887	0,018754	0,093403
Dnipropetrovs′ka	1	0,991909	0,876707	0,973535	0,869613	0,853504
Donets′ka	0,534427	0,346928	1	0,978811	0,185408	0,523103
Zhytomyrs′ka	0,17855	0,166661	0,656827	0,958539	0,019545	0,112414
Zakarpat·s′ka	0,134532	0,119822	0,678916	0,97702	0,010944	0,089237
Zaporiz′ka	0,413841	0,328208	0,843373	0,979649	0,114552	0,341919
Ivano-Frankivs′k	0,21307	0,211332	0,683133	0,980189	0,030761	0,142672
Kyyivs′ka	0,483397	1	0,833936	0,97425	0,403122	0,392742
Kirovohrads′ka	0,178652	0,198038	0,659036	0,958632	0,023317	0,112868
Luhans′ka	0,110819	0,093709	0,688153	0,990221	0,007146	0,075515
L′vivs′ka	0,439997	0,546428	0,732129	0,984955	0,176023	0,317288
Mykolayivs′ka	0,223948	0,294394	0,8	0,98189	0,052743	0,175914
Odes′ka	0,46356	0,482579	0,78253	0,974303	0,175056	0,353428
Poltavs′ka	0,445466	0,354134	0,759639	0,976266	0,119837	0,330362

End of Table 1

Rivnens′ka	0,163806	0,131584	0,753614	0,957309	0,016244	0,118176
Sums′ka	0,19315	0,16706	0,69257	0,989081	0,022348	0,132309
Ternopil′s′ka	0,123863	0,144863	0,601205	0,976178	0,010787	0,072693
Kharkivs′ka	0,580109	0,50552	0,742369	1	0,217705	0,430655
Khersons′ka	0,149694	0,139624	0,627108	0,980085	0,013107	0,092005
Khmel′nyts′k	0,190924	0,282378	0,676908	0,962305	0,036494	0,124366
Cherkas′ka	0,236253	0,197246	0,674699	0,967428	0,031441	0,154207
Chernivets′k	0,085992	0,078721	0,61245	0,980129	0,004146	0,051619
Chernihivs′ka	0,17177	0,160318	0,661647	0,959599	0,01822	0,10906

Source: author's calculations.

Note: [1]Excluding the temporarily occupied territory of the Autonomous Republic of Crimea, Sevastopol and the zone of the anti-terrorist operation.

The rating of the regions by the level of extensive and intensive development shows that the development of the regions in Ukraine mainly proceeds in the extensive path of development (Table 1). Almost all regions of Ukraine demonstrate a low level of intensive development. The integrated coefficient of intensive development for many territories is far from a maximum value; there are well distinguishable and huge discrepancies in the levels of the regions' intensive development. Such a gap between the natural and human resource potentials, on the one hand, and the level of the development of economic activity and its territorial organization within the regions, on the other hand, leads to investment unattractiveness of some territories.

The data mining methods have gained much popularity in assessing regional differentiation.

Data mining is the process of detecting and extracting from the original, previously unknown, data the non-trivial and practically valuable knowledge necessary for decision-making in various areas of human activity [20]. The fields of application of Data Mining include the following: analysis and management of risks; defining the distinctive features of customer behavior; the assessment of the creditworthiness of individuals and legal entities; the identification of emergencies; the segmentation of customers, products, services, and others.

Solving these tasks employs a rather large number of generic software products: SPSS, Statistica, SAS, Deductor, and others. This paper's analysis using the methods of Data Mining involved the application of the Deductor software, which includes the following analytical algorithms: neural networks, Kohonen's self-organizing maps, autocorrelation and regression, associative rules, decision trees.

For our study, we used the cluster analysis method, particularly, Kohonen's self-organizing maps as one of the most popular and frequently used methods for solving problems of the regional economy and assessing the differentiation of regions. In the context of our task, the result of cluster analysis was clusters of regions, united by indices of socio-economic development.

The self-organizing Kohonen map is an artificial neural network-based clustering technique that performs clustering and visualization tasks. This is a method of projecting a multidimensional space onto a space with a lower dimension (usually two-dimensional), and can also be used to solve problems of simulation modeling, forecasting, etc. [21].

Let us consider the mathematical substantiation of the problem.

Assume a set of objects $I = \{i_1, i_2, ..., i_n\}$, each of which can be described by the vector x_j $(j = 1, 2, ..., n)$ of parameters: $x_j = \{x_{j1}, x_{j2}, ..., x_{jm}\}$. It is required to construct a set of clusters C:

$$C=\{c_1, c_2, ..., c_k, ..., c_g\}, \quad (3)$$

where c_k is the cluster containing "similar" elements from the set I:

$$c_k = \{i_j, i_p | i_j \in I, i_p \in I \text{ ma } d(i_j, i_p) < \sigma\}, \quad (4)$$

where *s* is the magnitude defining the degree of proximity of objects to include them in one cluster;

$d(i_j, i_p)$ is the distance between objects, which for Kohonen's networks, is typically measured by the Euclid distance. If $d(i_j, i_p)$ is less than the *s* value then the objects belong to the same cluster. Otherwise, the objects are different and belong to different clusters [22].

A Kohonen's network is a single-layer network built from neurons under the rule of WTA (the winner takes all). Each neuron within the network is connected to all components of the input vector $x_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$, which is subject to clustering. The number of neurons equals the number of clusters. A neuron in the network is a linear weighted adder:

$$s_j = b_j + \sum_{i=1}^{m} w_{ij} x_i, \quad (5)$$

where *j* is the neuron's number;

i is the input's number (parameter);

s_j is the adder's output;

w_{ij} is the weight of the *i*-th parameter of the *j*-th neuron;

b_j is the threshold.

Each *j*-th neuron is an *m*-dimensional weight vector:

$$w_j = (w_{1j}, w_{2j}, ..., w_{mj}), \quad (6)$$

where *m* is the number of input vectors' components [22].

The training of Kohonen neutral network is based on the selection of weights and includes certain stages. After training the network, there is a visualization stage using Kohonen's maps.

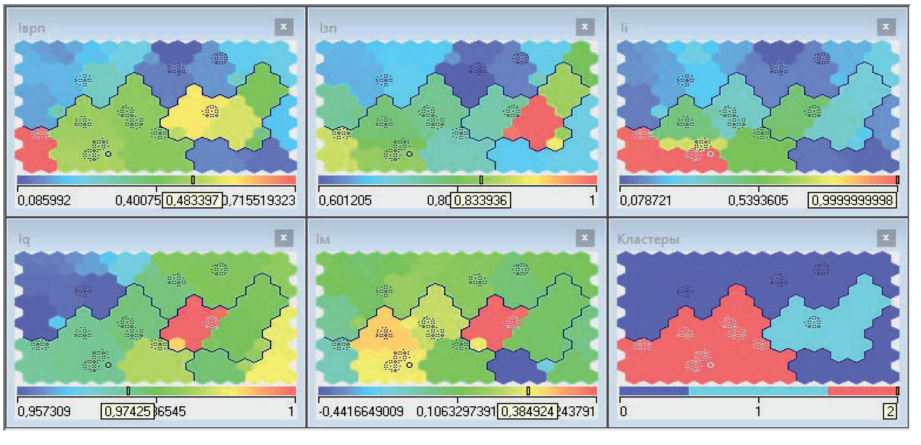
The data for our analysis are given in Table 1. For analysis, these particular indicators were chosen because their relationship coefficient was the largest. Before the construction of Kohonen maps, the data quality analysis was performed to detect and edit the surges and extreme data values. Editing employed the Deductor software package. After that, the optimal number of clusters was determined based on the results of the experimental study, which is 3. It is with this number of clusters that the most obvious division into groups could be provided. The results of clustering the regions of Ukraine in terms of their socio-economic development are shown in Fig. 2.

The distribution of oblasts by clusters using the Kohonen maps is given in Table2.

Let us consider the clustering of Ukraine's oblasts using the *k*-means algorithm, the result of which is given in Table 3.

The sum score based on the two clustering methods is given in Table 4.

The distribution of Ukraine's regions into three groups based on the ranking index is given in Table 5.



Source: author's calculations.

Fig. 2. The segmentation of Ukraine's oblasts in terms of their socio-economic development

Table 2

The clustering of Ukraine's oblasts using the Kohonen maps

Oblasts	Cluster No.	I_{GRP}	I_w	Ii	Iq	Im	Score sum
Kyyivs′ka	2	5	2	1	5	1	14
Dnipropetrovs′ka		1	1	1	5	5	13
L′vivs′ka		5	5	4	1	4	15
Odes′ka		5	4	5	4	1	19
Poltavs′ka		5	5	5	1	2	18
Kharkivs′ka	1	1	3	1	1	1	7
Donets′ka		1	1	3	3	3	11
Zaporiz′ka		3	2	3	3	3	14
Vinnyts′ka	0	1	3	2	5	1	12
Cherkas′ka		2	4	3	4	1	14
Ternopil′s′ka		5	5	4	3	1	18
Chernivets′k		5	5	5	2	1	18
Chernihivs′ka		3	4	4	5	1	17
Ivano-Frankivs′k		2	3	2	2	1	10
Mykolayivs′ka		2	1	1	2	1	7
Zhytomyrs′ka		3	4	3	5	1	16
Khmel′nyts′k		3	4	1	5	1	14
Kirovohrads′ka		3	4	3	5	1	16
Luhans′ka		5	3	5	1	1	15
Volyns′ka		4	4	3	5	2	18
Khersons′ka		4	5	4	2	2	17
Sums′ka		3	3	3	1	2	12
Rivnens′ka		3	2	4	5	2	16
Zakarpat·s′ka		4	4	5	3	5	21

Source: author's calculations.

Table 3

The clustering of Ukraine's oblasts using the *k*-means algorithm

Oblasts	Cluster No.	I_{GRP}	I_w	Ii	Iq	Im	Score sum
Kyyivs′ka	2	2	1	1	3	1	8
Dnipropetrovs′ka		1	1	1	3	5	11
L′vivs′ka		3	4	3	1	3	14
Ivano-Frankivs′k		5	5	5	2	3	20
Kirovohrads′ka		5	5	5	5	4	24
Luhans′ka		5	5	5	1	4	20
Volyns′ka		5	5	5	5	4	24
Vinnyts′ka		4	5	5	5	2	21
Kharkivs′ka	1	1	1	1	1	1	5
Sums′ka		4	2	4	2	4	16
Khersons′ka Khersons′ka		4	4	4	4	4	20
Ternopil′s′ka		4	4	4	4	4	20
Poltavs′ka	0	1	3	2	2	1	9
Cherkas′ka		4	4	4	3	2	17
Zaporiz′ka		2	3	2	1	3	11
Cherkas′ka		5	5	5	1	2	18
Chernihivs′ka		5	5	4	5	2	21
Odes′ka		1	3	1	2	1	8
Mykolayivs′ka a		4	4	3	1	2	14
Zhytomyrs′ka		4	5	4	5	2	20
Khmel′nyts′ka		4	5	3	4	2	18
Donets′ka		1	1	2	1	3	8
Rivnens′ka		5	5	5	5	3	23
Zakarpat·s′ka		5	5	5	1	5	21

Source: author's calculations.

Table 4

The sum score of Ukraine's oblasts

Oblasts	Score calculated by		The sum score	Ranking
	Kohonen maps	*k*-means algorithm		
Vinnytsʹka	12	21	33	11
Volynsʹka	18	24	42	1
Dnipropetrovsʹka	13	11	24	20
Donetsʹka	11	8	19	23
Zhytomyrsʹka	16	20	36	8
Zakarpat·sʹka	21	21	42	1
Zaporizʹka	14	11	25	19
Ivano-Frankivsʹk	10	20	30	14
Kyyivsʹka	14	8	22	21
Kirovohradsʹka	16	24	40	3
Luhansʹka	15	20	35	10
Lʹvivsʹka	15	14	29	15
Mykolayivsʹka	7	14	21	22
Odesʹka	19	8	27	17
Poltavsʹka	18	9	27	17
Rivnensʹka	16	24	40	3
Sumsʹka	12	16	28	16
Ternopilʹsʹka	18	20	38	5
Kharkivsʹka	7	5	12	24
Khersonsʹka	17	20	37	7
Khmelʹnytsʹk	14	18	32	12
Cherkasʹka	14	17	31	13
Chernivetsʹk	18	18	36	8
Chernihivsʹka	17	21	38	5

Source: author's calculations.

Table 5

The segmentation of Ukraine's oblasts

№ / rank	1 group	2 group	3 group
	24-17	16-9	8-1
1	Dnipropetrovsʹka	Vinnytsʹka	Vinnytsʹka
2	Donetsʹka	Ivano-Frankivsʹk	Zhytomyrsʹka
3	Zaporizʹka	Luhansʹka	Zakarpat·sʹka
4	Kyyivsʹka	Lʹvivsʹka	Kirovohradsʹka
5	Mykolayivsʹka	Sumsʹka	Rivnensʹka
6	Odesʹka	Khmelʹnytsʹka	Ternopilʹsʹka
7	Poltavsʹka	Cherkasʹka	Khersonsʹka
8	Kharkivsʹka		Chernivetsʹk
9			Chernihivsʹka

Source: author's calculations.

Conclusions. Fig. 2 shows that the division by all indicators is not completely homogeneous. If we consider, for example, division based on the I_{GRP} indicator (an integrated indicator of the gross regional product), it is the most uniform for cluster 2, which includes Kyiv, Dnipropetrovsk, Lviv, Odesa, and Poltava oblasts. For the oblasts of this cluster, the indicators of the gross regional product (I_{GRP}), investment (Ii), and wages (IAP) are the most significant (89–99.3 %) and they, in particular, made it possible for oblasts to be included in one cluster. The least significant indicator is Iq (the share of people of working age of the total population) because, for all oblasts of cluster 2, it varies within 0.97–0.98. It is also noticeable that the Dnepropetrovsk oblast is distinguished by the parameters I_{GRP}, Ii, and I_w, as evidenced by the uneven coloration for the corresponding indicators (Fig. 2).

If other clusters are analyzed, then, for cluster 0, whose strength is 16, the main significant indicators (93.5–98 %) are the indicators of the gross regional product (I_{GRP}), investment (Ii), and wages (I_w). For cluster 0, of strength 3, of significance are the indicators of the gross regional product (I_{GRP}), wages (I_w), and the share of people of working age in the total population (Iq).

Given this, we can argue that the indicators of the gross regional product (I_{GRP}), wages (I_w), and investment (Ii) are crucial for clustering the regions of Ukraine in terms of their socio-economic development. This is confirmed by the correlation analysis carried out using the Deductor software, in which the values of the above factors fluctuate within 0.65–0.85.

However, the analysis of Table 2 and Fig. 2 makes it clear that many factors affect the way an oblast is included in a cluster. For example, Dnipropetrovska oblast, based on the I_{GRP} indicator, is close to cluster 0 (Fig. 2) but, based on other indicators, it included in cluster 2. Therefore, it is expedient to apply additional methods for a more accurate clustering of Ukraine's oblasts by the level of socio-economic development and the final ranking.

Let us consider the clustering of Ukraine's oblasts using a *k*-means algorithm, the result of which is given in Table 3. Based on the clustering results, given in Table 3, one can note that the indicators of the gross regional product (I_{GRP}), wages (I_w), and investment (Ii) were the most crucial to be included in cluster 1. Their significance exceeded 98 %. However, for cluster 0, the list of defining indicators somewhat changed. Namely: the indicator of the gross regional product (I_{GRP}), investment (Ii), and the share of people in working age in total population (Iq), whose weight exceeds 90 %. For cluster 2, the defining indicators are the gross regional product (I_{GRP}), wages (I_w), and the share of people in working age in total population (Iq).

The sum score based on the two methods of clustering is given in Table 4. The distribution of Ukraine's oblasts into three groups based on the ranking index is given in Table 5. Thus, group 2 includes those oblasts that are characterized by high (compared to the rest) indicators of I_{GRP}, I_w, and Ii. The exclusion is Kherson oblast, and Odesa oblast is more appropriate. These oblasts correspond to a sufficient, stable level of life. This opinion is shared by V. Ignatenko [23], who divided the regions of Ukraine into clusters based on the level of the integrated regional index of human development, while Dnipropetrovska, Zaporizka, Kyyivska, Odeska, and Lvivska oblasts are categorized as stable.

The defining indicators for the first group are I_{GRP}, I_w, and Iq.

The third group is characterized by the low level of wage, gross regional product. Most of them are characterized by a low level of production capacity [24] but have sufficient capacity to form the budget potential [23].

Thus, the estimation of the country's regions based on the level of their socio-economic development testifies to the dominance of extensive factors in the development of most regions in Ukraine. Common areas of the policy, conducted in the labor market, for all groups of regions are the measures to conduct an active policy (promoting self-employment and small businesses; the creation of new jobs; vocational training and retraining of unemployed people; public works; improvement of employment services, etc.).

Finally, it should be mentioned that the idea of clustering regions of a country based on the level of their socio-economic development using data mining methods is not new. However, our study provides an opportunity to assess regional labor markets in terms of the socio-economic development of regions based on several methods: Kohonen maps and *k*-means. The latter makes it possible to reduce inaccuracies and disadvantages of a single particular method.

References

1. Chadi, A., Pinto, M., Schulze, G. (2019) Young, gifted and lazy? The role of ability and labor market prospects in student effort decisions. *Economics of Education Review*, Volume 72, pp. 66-79. doi: 10.1016/j.regsciurbeco.2018.12.003

2. Schwab, K. (2017). The Fourth Industrial Revolution, Crown Business. New York, ISBN 9781524758868.

3. Acemoglu, D., Restrepo, P. (2017). *Robots and jobs: evidence from US labor market.* NBER working paper.

4. Chinoracký, R., Čorejová, T. (2019) Impact of Digital Technologies on Labor Market and the Transport Sector. *Transportation Research Procedia*, 40, pp. 994-1001. doi: 10.1016/j.trpro.2019.07.139

5. Mäkelä, E. (2017). The effect of mass influx on labor markets: Portuguese 1974 evidence revisited. *European Economic Review, Volume 98,* 240-263.

6. Ciani, E., David, F., Blasio, G. (2019) Local responses to labor demand shocks: A Re-assessment of the case of Italy. *Regional Science and Urban Economics*, Volume 75, pp. 1-21. doi: 10.1016/j.regsciurbeco.2018.12.003

7. Goulas, E., Zervoyianni, A. (2018) Active labour-market policies and output growth: Is there a causal relationship. *Economic Modelling*, Volume 73, pp. 1-14. doi: 10.1016/j.econmod.2017.11.019

8. Tejada, M.M. (2017) Dual labor markets and labor protection in an estimated search and matching model. *Labour Economics*, Volume 46, pp. 26-46. doi: 10.1016/j.labeco.2017.03.002

9. Dengler, K., Matthes, B. (2018) The impacts of digital transformation on the labour market: Substitution potentials of occupations in Germany. *Technological Forecasting and Social Change*, Volume 137, pp. 304-316. doi: 10.1016/j.techfore.2018.09.024

10. Skills, not job titles, are the new metric for the labour market (2019). Official website of the World Economic Forum. Available at: https://www.weforum.org/agenda/2019/07/skills-not-job-titles-are-the-new-metric-for-the-labour-market/ (Accessed 16 September 2019).

11. Pedley, D., McHenry, D., Motha, H. and Shah, J.N. (2018). Understanding the UK cyber security skills labour market. *Research report for the Department for Digital, Culture, Media and Sport*. *Ipsos MORI.* London, 76.

12. Ministry of Economy and Sustainable Development of Georgia (2017). *Labour Market Analysis of Georgia*, 25.

13. Böckermana, P., Haapanen, M., Jepsen, C. (2019) Back to school: Labor-market returns to higher vocational schooling. *Labour Economics*, Volume 61, 101758. doi: 10.1016/j.labeco.2019.101758

14. Selwaness, I., Zaki, C. (2019) On the interaction between exports and labor market regulation: Evidence from the MENA countries. *The Quarterly Review of Economics and Finance,* Volume 73, pp. 24-33. doi: 10.1016/j.qref.2018.05.011

15. Prescott, J.J., Pyle, B. (2019) Identifying the impact of labor market opportunities on criminal behavior. *International Review of Law and Economics*, Volume 59, pp. 65-81. doi: 10.1016/j.irle.2019.04.001

16. Xu, X., Li, D.D., Zhao, M. (2018) "Made in China" matters: Integration of the global labor market and the global labor share decline. *China Economic Review*, Volume 52, pp. 16-29. doi: 10.1016/j.chieco.2018.05.008

17. Savić, M., Zubović, J. (2015) Comparative Analysis of Labour Markets in South East Europe. *Procedia Economics and Finance*, 22, pp. 388-397. doi: 10.1016/S2212-5671(15)00309-3

18. Shigapova, D., Valiullin, M., Yrieva, O., Safina, L. (2015) The Methods of Prediction of Demand on The Labor Market. *Procedia Economics and Finance*, 23, pp. 1476-1479. doi: 10.1016/S2212-5671(15)00477-3

19. Svetunkov, S.G., Zagranovskaya, A.V., Svetunkov, I.S. (2012) *Kompleksnoznachnyy analiz i modelirovaniye neravnomernosti sotsial'no-ekonomicheskogo razvitiya regionov Rossii* [A comprehensive analysis and modeling of the uneven socio-economic development of the Russian regions]. *S.-Pb*, 129 p.

20. Paklin, N.B., Oreshkov, V.I. (2013). *Biznes-analitika: ot dannykh k znaniyam: uch. posobiye* [Business analytics: from data to knowledge: textbook]. *S.-Pb. Peter*, 704 p.

21. Anisimova, E.S. (2014) *Samoorganizuyushchiyesya karty Kokhonena v zadachakh klasterizatsii* [Kohonen self-organizing maps in clustering problems]. *Aktual'nyye problemy gumanitarnykh i yestestvennykh nauk,* 9(68), pp. 13-16.

22. Gorbachenko, V.I. *Seti i karty Kokhonena* [Kohonen Networks and Maps]. Available at: http://gorbachenko.self-organization.ru/articles/Self-organizing_map.pdf (Accessed 5 January 2021)

23. Ihnatenko, V.Yu. (2012) *Udoskonalennya mekhanizmu formuvannya byudzhetnoho potentsialu rehioniv na osnovi metodu klasteryzatsiyi* [Improving the mechanism of forming the budget potential of regions based on the method of clustering]. *Naukovyy visnyk Poltavs'koho universytetu ekonomiky i torhivli*, 2 (53), pp. 70-75.

24. Kravets′, T.V., Verhay, T.I. (2017) *Otsinyuvannya rivnya vyrobnychoho potentsialu rehioniv Ukrayiny z vykorystannyam neyronnykh merezh* [Assessment of the level of production potential of the regions of Ukraine using neural networks]. *BiznesInform,* 11, pp. 112-119.

ASSESSING THE REGIONAL LABOR MARKET BY USING DATA MINING METHOD: WAYS OF EFFECTIVE FUNCTIONING

Larysa D. Harmider, SHEI Ukrainian State Chemical Technology University, Dnipro (Ukraine).

E-mail: garm@ukr.net

Svitlana O. Fedulova, Alfred Nobel University, Dnipro (Ukraine).

E-mail: sveta_fedulova@ukr.net

Yuliia M. Bartashevska, Alfred Nobel University, Dnipro (Ukraine).

E-mail: bartashevska@duan.edu.ua

Vitalina V. Komirna, European University Servant of God Robert Schuman, Radom (Poland).

E-mail: v.komirna@gmail.com

DOI: 10.32342/2074-5354-2022-2-57-3

Key words: labor market, region, data mining methods, indicators, socio-economic development.

As a result of the uneven development of certain territories, it is more feasible and effective to tackle the practical issues of labor market regulation at the regional level. This ensures sufficient regulation of the system. Since it is necessary to properly account for the regional differences in practice, it is required that these issues be methodologically justified. Therefore, the aim of this paper is to investigate regional labor markets based on indicators of the socio-economic development of regions using the data mining methods.

The current study has clustered regions of Ukraine on the basis of the level of their socio-economic development using data mining methods, in particular Kohonen maps and the k-means methods. One of the most critical stages in the assessment of Ukraine's regions in terms of socio-economic development by using data mining methods is to determine the information base, criteria of evaluation, and a list of estimates. The data mining methods have gained much popularity in the assessing regional differentiation.

The conducted analysis based on data mining methods included the use of the Deductor software, which includes the following analytical algorithms: neural networks, Kohonen's self-organizing maps, autocorrelation and regression, associative rules, decision trees. For our study, we used the cluster

analysis method based on Kohonen's self-organizing maps as one of the most popular and frequently used methods for solving problems of the regional economy and assessing the differentiation of regions. In the context of our task, the result of cluster analysis is clusters of regions, united by indices of socio-economic development. The main aspects of the socio-economic and demographic development of the regions are characterized by a set of statistical indicators related to four blocks of key factors: 1. Assessment of the demographic situation in a region. 2. Assessment of the social situation in a region. 3. Assessment of the economic situation in a region. 4. Assessment of the organizational environment in a region.

The study, by no means, claims to detect all the dependences in the labor market related to all the above-mentioned factors. Based on public data, given in the statistical yearbook "Ukraine in Figures" (2020), by using mathematical methods (correlation-regression and cluster analysis), we obtained two groups of factors that characterize different aspects of the socio-economic and demographic development.

The ranking of the regions by the level of extensive and intensive development shows that the development of the regions in Ukraine mainly proceeds in the extensive path of development. Almost all regions of Ukraine demonstrate a low level of intensive development. The integrated coefficient of intensive development for many territories is far from a maximum value; there are well distinguishable and huge discrepancies in the levels of the regions' intensive development. Such a gap between the natural and human resource potentials, on the one hand, and the level of the development of economic activity and its territorial organization within the regions, on the other hand, leads to investment unattractiveness of some territories.

Thus, the estimation of the country's regions based on the level of their socio-economic development testifies to the dominance of extensive factors in the development of most regions in Ukraine. Common areas of the policy, conducted in the labor market, for all groups of regions are the measures to conduct an active policy (promoting self-employment and small businesses; the creation of new jobs; vocational training and retraining of unemployed people; public works; improvement of employment services, etc.).

Одержано 28.09.2022.