Using Data Mining Approaches to Study Youth Readiness for Industry 4.0

Yuliia Bartashevska, Halyna Ryzhkova

Alfred Nobel University, Ukraine
bartashevska@duan.edu.ua, g.ryzhkova@duan.edu.ua

Abstract

One of the most pressing and most discussed issues in the modern world is the impact of Industry 4.0 on the labor market. These technological changes are not only related to automation, digitalization and artificial intelligence, they are capable of changing the foundations of the economy, namely: the methods of production and business organization, lead to significant changes in the labor market, and as a result affect the distribution of jobs and the successful career of any person.

To successfully adapt to these changes, it is necessary to invest in education and vocational training so that young people can acquire the necessary skills and adapt to new technologies. In addition, it is important to develop policies and support measures that will help workers whose professions are becoming obsolete due to automation to retrain and find new employment opportunities.

The study conducted a comparative analysis of the level of readiness of young people for employment in the conditions of Industry 4.0 using the example of two countries - Poland and Ukraine. The findings indicate that the problem of youth employment is caused not only by technological challenges, but also by the insufficient use of educational opportunities and the slow adaptation of institutions to digital changes.

Keywords: Industry 4.0, automation, digitalization, artificial intelligence, youth employment, education, professional adaptation.

1. Introduction.

Despite potential threats, automation, digitalization, and artificial intelligence create new opportunities for development.

Manifestations of Industry 4.0 can increase the efficiency of production and information processing. Robotization of production allows you to reduce the time for production processes, reduce the likelihood of errors, and improve product quality [1, 2]. Digitalization of business processes simplifies and speeds up the performance of tasks, reduces bureaucratic burdens, and increases the availability of information [3]. Artificial intelligence makes it possible to automate many tasks that previously required human intervention, such as data analysis, trend forecasting, and decision-making.

However, these technological changes also create challenges and threats for the labor market. Automation and robotization can lead to a decrease in demand for certain types of labor, especially for routine and low-skilled jobs. A number of professions may become obsolete or lose their demand due to automation. This can lead to increased unemployment and job instability, especially among low-skilled workers [4, 5], for example, young people [6].

New technologies require new skills and competencies, which can lead to the emergence of new types of jobs and a demand for specialists in information technology, engineering, data analytics and other related fields [7]. In addition, automation can make some types of work safer and less monotonous, freeing people from routine tasks and allowing them to focus on more creative and strategic aspects of their work [8].

Automation, digitalization and artificial intelligence have a profound and multifaceted impact on the labor market, posing certain challenges and threats, but they also create new opportunities for development and growth [9]. Effective management of these changes requires joint efforts from governments, businesses, educational institutions and society as a whole.

Digital technologies have the greatest impact on society and labor relations, significantly transforming the issues of employment and unemployment of any person. At the same time, structural shifts in the modern economy, changes in the content and organization of work, the blurring of the boundaries between the employed and the unemployed, dynamic changes in the structure of the working population have an ambiguous and contradictory impact on youth employment, deepening the problem of youth unemployment in the context of Industry 4.0.

2. Methodology.

The study of the readiness of young people in Ukraine and Poland was conducted in 2022 using an integrated approach that combines quantitative and qualitative analysis methods. The main empirical tool was a survey of students using the CAWI (Computer-Assisted Web Interviewing) method, which allowed collecting structured data from 1,155 respondents from both countries (45% were respondents from Ukraine, 55% from Poland), aged 18-26. The survey was conducted using an author's questionnaire designed to self-assess students' awareness, attitude and level of preparedness for the conditions of the digital economy and Industry 4.0, where the answers were graded from "absence" of skills to "very high" level of proficiency. The representativeness of the sample was ensured by adhering to the standards of sociological research: the maximum error does not exceed 4%, the confidence level is 98%. Additionally, a content analysis of relevant international studies, policy documents, and statistical reports was conducted to identify global trends and risks associated with the impact of digital technologies on the labor market.

3. Discussion.

To process the results of the study, the Delta Model from the global McKinsey Institute was chosen [10]. The model is the result of a survey of 18,000 people in 15 countries and allows us to determine the basic skills of the future for educating citizens in a digital environment. According to McKinsey, soft skills and hard skills should be divided into 4 categories: cognitive, digital, interpersonal and self-leadership. Each group has 3-4 subgroups, which in turn contain certain skills. A total of 56 skills in 13 groups and 4 categories [10].

Let us divide the results of our study into 4 categories according to the Delta Model (Table 1).

Table 1.

Distribution of research skills by Delta Model categories

Category	Group	Skill
Cognitive skills	Critical thinking	Structured problem solving
		Logical reasoning
		Understanding biases and seeking relevant information
	Planning and ways of working	Agile thinking
	Mental flexibility	Creativity and imagination
		Adopting to different contexts
		Desire and ability to learn
Interpersonal skills	Developing relationships	Communication skills
	Teamwork effectiveness	Resolving conflicts
		Collaboration
Self-leadership	Self-awareness and self-management	Understanding own emotions and triggers
	Goals achievement	Result orientation
		Self-improvement
Digital skills	Digital fluency and citizenship	Digital literacy, AI
		Digital literacy, cloud computing
		Digital literacy, automation, and robotics
		Digital literacy, AR/ VR
	Software use and development	Programming literacy
		Data analytics, Big Data
		Algorithmic thinking
	Understanding digital systems	IoT
		Cybersecurity literacy

According to Table 1 and the results of our survey, we have 22 skills, divided into 4 categories and 10 groups.

4.Results.

To determine the real distribution of competencies in the sample and identify hidden skill groups with similar proficiency profiles, we will cluster the survey results distributed according to Table 1 using the k-means method.

The algorithm is one of the most common, simple, but effective, the principle of which is to minimize the distance between each object in the cluster and the center of the cluster by calculating the Euclidean distance [11]. K-means works by iteratively distributing objects (in our case, skills) into clusters, minimizing the sum of the squares of the distances from each object to the centroid of its cluster. Another advantage of the method is fast convergence, but it requires an initial definition of clusters. It is the careless determination of their number that can lead to inaccuracy of the method [12].

To determine the number of clusters according to [13], you can use: the elbow method, gap statistics, the silhouette method, and the canopy method, each of which has its own advantages and disadvantages. We will use the elbow method, which allows us to visually (based on the graph) find the point where adding a new cluster no longer leads to a significant improvement in the quality of the model. For our study, k=3 (Fig. 1) is optimal, as it provides a clear division of skills into groups.

Also, before performing clustering using the k-means method, preliminary cleaning and data reduction were performed, as well as their normalization. Each skill was converted from absolute values (number of people) to a percentage distribution for each of the five proficiency levels. This approach ensures that each skill has the same weight in the clustering process, regardless of the total number of respondents who answered a specific question.

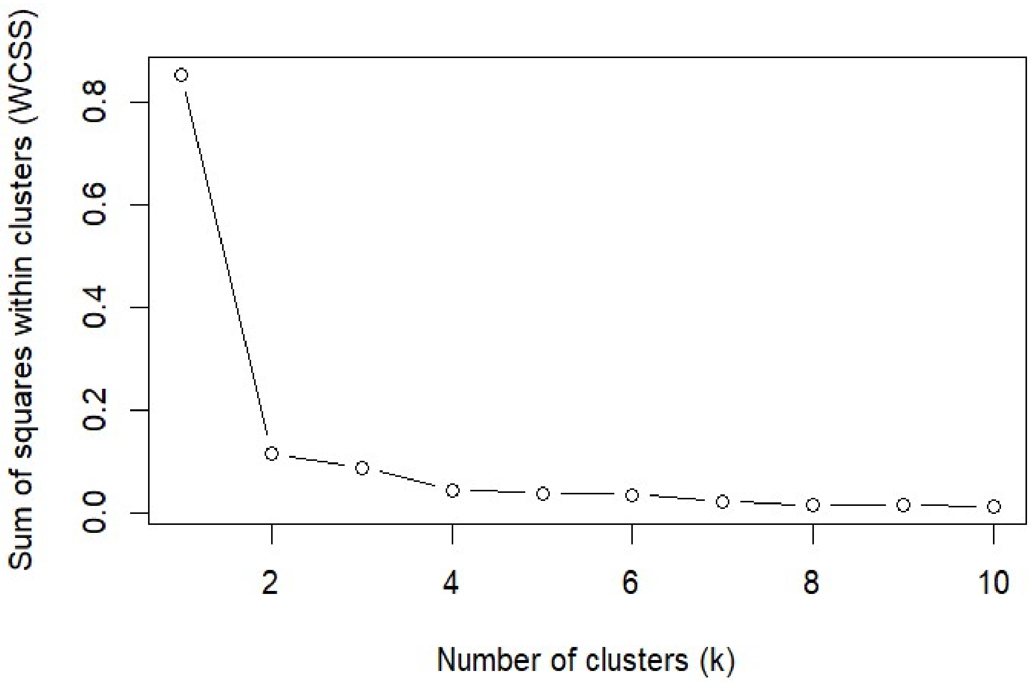

Fig. 1. Determine the count of clusters by Elbow Method

The results of the cluster analysis revealed three clearly defined skill groups, each characterized by a unique proficiency profile. Each cluster was given a name reflecting its main characteristics, namely: Technological and Analytical Competencies (1), Socio-Communicative Competencies (Basic) (2), Digital Engineering and Programming Competencies (3) (Fig.2).

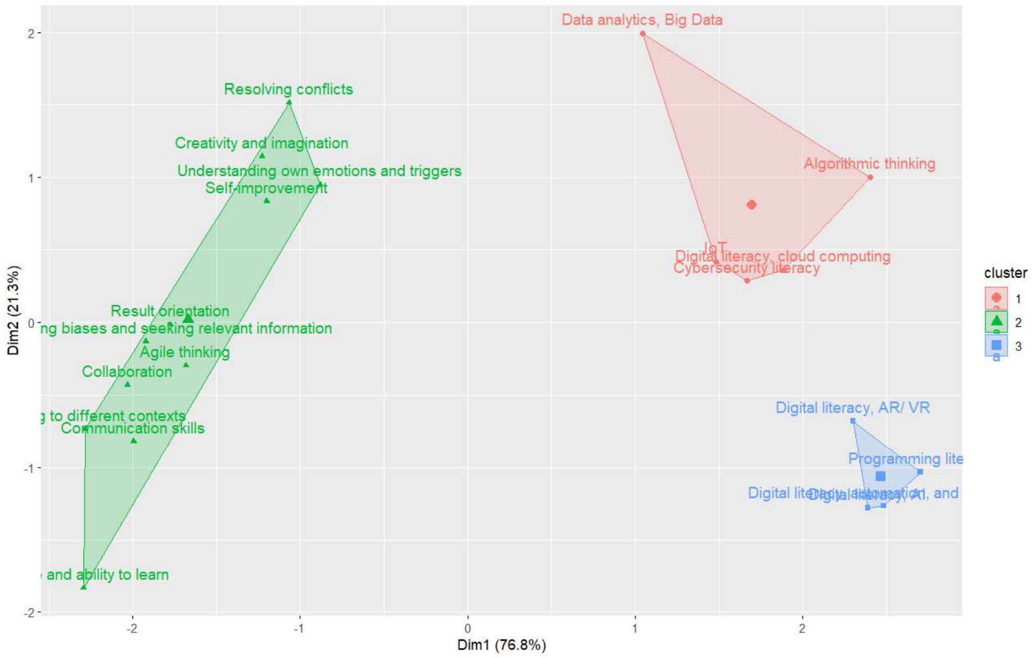

Fig. 2. Skill clustering by the K-means method

The cluster *Social and Communicative Competencies (basic) (2)* is the most numerous and includes cognitive and interpersonal skills. The cluster is characterized by a high concentration of responses in the categories "High" and "Average" (37.6% and 32.5%, respectively). This indicates that the skills combined in this cluster are widely developed for most applicants. These include such competencies as: Agile Thinking, Structured Problem Solving, Logical Reasoning, Communication skills, Collaboration, etc. The main goal for the skills of this cluster is not so much in their intensive development, but in their maintenance and effective use for solving problems and establishing cooperation.

The cluster *Technological and Analytical Competencies (1)* in terms of the number of skills in it almost coincides with the cluster *Digital Engineering and Programming Competencies (3)*, but the latter contains more highly specialized skills: Digital literacy, AI; Digital literacy, automation, and robotics; VR/AR; Programming literacy. Skills of the cluster Technological and Analytical Competencies (1), namely: Digital literacy, cloud computing; Data analytics, Big Data; Algorithmic thinking; Internet of Things (IoT); Cybersecurity literacy have a higher percentage of “Very high” (12%) and “High” (29.8% of respondents) compared to Cluster 3, and less “None” (7.5% of respondents). Skills of cluster 3 are relatively new (except for programming), highly specialized, they have the highest percentage of “None” (18.3%) and “Low” (26.8% of respondents), and the lowest percentage of “Very high” (8.8%). This gap can be explained by the fact that most of these skills began to be implemented in 2022, and the surveyed students do not yet have a good command of them. The *Digital Engineering and Programming Competencies Cluster (3)* identifies the needs of the future, signals the largest educational gap, and the need for strategic investments in training and the development of the necessary skills market.

5. Conclusion.

Based on the conducted study of the readiness of students in Ukraine and Poland to work in the conditions of Industry 4.0, and the skills they have acquired, we can draw the following conclusions:

- the strengths are the Cognitive Skills, Interpersonal Skills and Self-Leadership groups, which have a high level of self-esteem among students. These are the skills that need to be combined/integrated with technical ones;
- the digital skills group is divided into two clusters with almost the same number of components. Cluster 1 should include skills that are already necessary for the current market, and cluster 3 consists of promising skills that are needed by the market and those that require investment. This is

where the biggest gap in competencies lies. In general, digital skills in both clusters require development and comprehensive programs to improve them, as more than 25% of respondents have a "None" proficiency level and almost 45% have a "Low" proficiency level.

Certificate courses and/or mandatory modules on cyber literacy development and LLM application can be used to develop digital competencies. Also, workshops and hackathons on programming and robotics, specialized lectures and internships at leading technology companies during training should change the employment situation of young people for the better.

References:

1. Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. Journal of Economic Perspectives, 33(2), 31–50.
2. Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. Journal of Political Economy, 128(6), 2188–2244.
3. Дашко, І., Череп, О., & Михайліченко, Л. (2024). Розвиток штучного інтелекту: переваги та недоліки. Економіка та суспільство, (67). https://doi.org/10.32782/2524-0072/2024-67-31
4. Humlum, A. (2019). Robot adoption and labor market dynamics [Working paper]. Princeton University.
5. Im, Z. J. (2020). Automation risk and support for welfare policies: How does the threat of unemployment affect demanding active labour market policy support? Journal of International and Comparative Social Policy, 37(1), 76–91.
6. Im, Z. J., Mayer, N., Palier, B., & Rovny, J. (2019). The "losers of automation": A reservoir of votes for the radical right? Research & Politics, 6(1).
7. Inglehart, R., & Norris, P. (2019). Cultural backlash: Trump, Brexit, and the rise of authoritarian populism. Cambridge University Press.
8. Marenco, M., & Seidl, T. (2021). The discursive construction of digitalization: A comparative analysis of national discourses on the digital future of work. European Political Science Review, 13, 391–409.
9. Syed, R., et al. (2020). Robotic process automation: Contemporary themes and challenges. Computers in Industry, 115, 103162.
10. Dondi, M., Klier, J., Panier, F. and Schubert, J. (2021). Defining the skills citizens will need in the future world of work. URL: https://www.mckinsey.com/industries/public-sector/our-insights/defining-the-skills-citizens-will-need-in-the-future-world-of-work, last accessed 2025/04/24.
11. Indra, H., M Alfin K., Triyani, Mutia Nur E., Renny (2025). A more precise elbow method for optimum k-means clustering. URL: https://arxiv.org/pdf/2502.00851, last accessed 2025/09/26.
12. X Li, L Yu, L Hang, and X Tang (2017). The parallel implementation and application of an improved k-means algorithm. J. Univ. Electron. Sci. Technol, 46:61–68, 2017.
13. Chunhui Yuan and Haitao Yang. Research on k-value selection method of k-means clustering algorithm. J, 2:226–235, 6 2019.