



Labor Resources of the Periphery of the Digital Economy: Identifying Signs Using a Machine Learning Model

Dmitriy V. Chuprakov , Anastasia A. Sozinova ,
Nadezhda K. Savelyeva , Victoria A. Saidakova ,
and Alexander S. Astakhin

Abstract

The research aims to check the hypothesis and identify signs of labor resources in the periphery of the digital economy using a machine learning model. The research is based on the analysis of a vacancy from the HeadHunter recruitment platform in terms of the requirements stated by the employer for the applicant and the identification of signs of labor resources using a machine learning model. The main methods in the study are general scientific methods, semantic text analysis methods, machine learning methods, principal component methods, statistical methods, and others. Based on the proposed machine learning model, the authors defined the characteristics of labor resources in the periphery of the digital economy in the context of aggregated categories of economic activity. The profiles of the required employees are individual for each type of economic activity due to the specific features of the activities of enterprises in the periphery of the digital economy. The authors identified factors that influence the decision-making process in the periphery of the digital economy. The authors emphasize that the signs of labor resources on the periphery of the digital economy can be differentiated depending on the industry of the enterprise.

Keywords

Labor resources · Machine learning model · Core of the digital economy · Periphery of the digital economy · Vacancies · Open data

JEL Classification

C45 · C52 · C53 · C54 · E00 · O30 · O33 · J24

1 Introduction

Digitalization is the basis for the development of the digital economy. More enterprises in various fields of activity use it annually. As a result, they put forward new digital requirements for labor resources.

In Russia, this process is closely related to achieving national goals and strategic objectives of state development within the framework of the national program “Digital economy of the Russian Federation.” Its implementation helps increase the level of digital processes in the economic and social sectors. This fact underlines the importance of studying the development of digitalization in the Russian economy. To take a place in the global digital economy, Russia needs to determine the characteristics of a Russian worker who meets the requirements of advanced digital technologies. Digitalization significantly affects the labor market in terms of labor requirements and shapes the competitiveness of enterprises (Sozinova et al., 2023d). Competitive employees contribute to improving the quality of products (services) through the skills of managing high-tech processes as part of adaptation to new working conditions (Sozinova et al., 2023a, 2023b). The development of

D. V. Chuprakov · A. A. Sozinova (✉) · N. K. Savelyeva ·
V. A. Saidakova
Vyatka State University, Kirov, Russia
e-mail: aa_sozinova@vyatsu.ru

D. V. Chuprakov
e-mail: usr10381@vyatsu.ru

N. K. Savelyeva
e-mail: nk_savelyeva@vyatsu.ru

V. A. Saidakova
e-mail: usr21787@vyatsu.ru

A. S. Astakhin
Belgorod State National Research University, Belgorod, Russia
e-mail: aastaxin@yandex.ru

digital skills is also possible within an organization through corporate social responsibility (Ashmarina et al., 2023).

However, not every sector of the economy has the necessary level of penetration of digital technologies into the activities of enterprises. Therefore, the characteristics of labor resources operating within the digital economy are not comparable. The development of digital technologies may require the acquisition of skills in each industry in accordance with the changes occurring in the era of the digital economy, including through the introduction of new educational programs (Shaengchart & KraiwJanit, 2022; Speshilova et al., 2023).

Belousov and Timofeeva (2019), Chukov and Zolotov (2021), Kaibicheva et al. (2020), Lumpuva (2018), Siqueira and Fontenelle (2023), Tagarova (2020), Zolotov (2022), and others studied the issues of identifying the periphery of the digital economy. They adapted the world-system analysis of Wallerstein (2006) to study the level of digitalization of national economies. The periphery of the digital economy consists of industries where the use of new digital technologies is sporadic or absent.

Economic development largely correlates with the successful activities of enterprises and their labor resources. Labor resources and the emerging demand for them in the labor market depend on the economic transformation. In other words, if the economy is focused on ecology, then the demand for green professions will grow (Lin and Zhou, 2024; Sozinova et al., 2023c). The same is true with the transition of the economy to digital priorities. To increase labor productivity and the level of digitalization, priority will be given to digital personnel who are initially ready to work in Industry 4.0 and have the appropriate skills and education (Tolmachev et al., 2023).

The interpretation of the labor resources of the periphery in scientific articles is not identical. Ch. Leadbeater (Radaev, 2008) defines them as part-time workers or self-employed. Bogomolova (2012) points out that the labor resources of the peripheral segment include those with limited working capacity and those with age restrictions. Other scientific researchers interpret the labor resources of the periphery in the form of people who have part-time and seasonal work, while their activities are quite labor-intensive (Kazantseva et al., 2020). This research proposes to consider labor resources from the perspective of the periphery of the digital economy. Employees work in economic sectors where enterprises either do not use digital technologies or use them at a minimal level.

The research hypothesis is that the characteristics of labor resources in the periphery of the digital economy will differ depending on the enterprise's industry. The research aims to check this hypothesis and identify signs of labor resources in the periphery of the digital economy using a machine learning model.

2 Methodology

To determine the characteristics of the labor resources of enterprises in the periphery of the digital economy, the authors studied the needs of employers based on job descriptions. The source of primary data was the HeadHunter recruiting platform (HeadHunter, n.d.), used to collect more than 130.2 thousand vacancies represented by 22.4 thousand enterprises. Eighteen enlarged types of economic activity were studied.

The methodological basis for formulating the requirements was a semantic analysis of the text. Then, the authors identified key and significant properties of the applicant for the enterprise as an employer. The texts of the requirements were divided into phrases based on the punctuation marks (“.” “;” “;” “;”) as employers approached the formatting of texts individually. As a result, the authors had 76,226 different phrases where there were no duplicates.

This text analysis was carried out using machine learning methods through the multilingual, freely distributed pre-trained embedding neural network model E5-Large (Wang et al., 2022). It transforms the source text into 1024-dimensional real-valued vectors. The distance between them decreases in the case of semantic similarity of texts. The choice of this model is based on studies of the relevance of responses from modern multilingual embedding neural network models of semantic text analysis (Wang, 2022) on the multilingual test sample Mr. TYDI (Zhang et al., 2021). The basis for assessing the quality of the model is the average reverse rank metric MRR@10. The closer its value is to 100, the closer to the target vector generated by the model are the vectors of phrases that are close in meaning. The MRR@10 value for the E5-large model based on texts in Russian is 65.8. This model also generates values in a very narrow cone when the distance between normalized vectors is almost always < 0.3 .

When reducing the dimensionality of the vectors, the authors used principal component analysis to reduce the influence of model properties and estimate the amount of the explained variance (Saydakova et al., 2024).

Thus, the dimension of vectors representing phrases is reduced at 99% of the explained variance from 1024 to 592, and at 60% to 69. Reducing the share of explained variance below 60% is not justified due to the increased rate of information loss.

In this case, the principal component method actually projects a set of vectors onto a subspace of lower dimension. Therefore, it leads to an increase in the distance between them.

Clustering of vector representations of semantically similar requirements is formed using the k-means method. By summing all vectors of a cluster and dividing them by the appropriate number of vectors, the centroid of the cluster is obtained. Cluster numbers are announced using new terms

in the requirements specification. Thus, the text expressions in the cluster are replaced by their number. As a result, average cluster requirements for the applicant are created.

Using formula 1, the authors note that the number of clusters is proportional to three components: the number of consolidated categories of types of economic activity, the number of levels of digitalization, and the number of high-lighted phrases.

$$n_{\text{cluster}} = \left[\frac{p}{100} \cdot \frac{n_{\text{elements}}}{n_{\text{category}} \cdot n_{\text{level}}} \right] \cdot n_{\text{category}} \cdot n_{\text{level}} \quad (1)$$

where:

n_{cluster} is the number of clusters;

n_{elements} is the number of highlighted phrases—requirements for candidates specified in vacancies (76,226);

n_{category} is the number of consolidated categories of economic activities (18);

n_{level} is the number of digitalization levels p —proportion of the number of clusters from the number in percent (4).

While explaining formula 1, the authors proposed four levels of digitalization. The zero level assumes that the requirements for the employee do not contain digital characteristics. The first level requirements are characterized by interaction with Internet technologies and basic office programs. The first two levels of digitalization represent the periphery of the digital economy. Production management and business process automation are signs of the requirements of the second level of digitalization, characterizing the semi-periphery of the digital economy. In this research, the periphery and semi-periphery of the digital economy are combined. The core of the digital economy edge includes requirements related to information technology development tools.

Then, each cluster label is considered a new term, replacing all phrases included in the cluster. To identify the most significant of them, the authors used a statistical measure to assess the significance of the term TF-IDF (formulas 2 and 3).

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D) \quad (2)$$

$$\text{tf}(t, d) = \frac{n_d(t)}{N_d}, \quad \text{idf}(t, D) = \ln \frac{D}{D_t} \quad (3)$$

where:

$\text{tf}(t, d)$ is the frequency of word t in document d ;

$\text{idf}(t, D)$ is the logarithm of the inverse frequency of occurrences of documents containing the word t in the collection;

$n_d(t)$ is the number of occurrences of the word t in the document d ;

N_d is the number of words in document d ;

D is the total number of documents in the collection;

D_t is the number of documents in the collection containing the word t .

The greater the resulting significance, the more often the term occurs within the studied document and less often in others.

The authors extracted teb cluster names that have the maximum TF-IDF measure value. This is how a list of clusters is formed that largely characterize the average needs of enterprises from a given enlarged category of types of economic activity with a given level of digitalization.

3 Results

As a result of applying the above stages of implementing a machine learning model, the authors obtained clusters of labor resource attributes. For example, an employer requirement such as “knowledge of sales techniques” includes not only this semantic unit but also requirements for managing sales and online sales; knowledge is also interpreted as “understanding of sales technology.”

The cluster “work process organization skills” includes the ability to work independently, self-organization, and process optimization. Given semantic proximity, this cluster also includes “the ability to organize work on construction sites,” “the ability to distribute the workload of warehouse employees,” and “skills for maintaining office life,” which already characterize the workplace specifics. Even “Microsoft products organizational skills” were included in this cluster due to the fact that the author of the vacancy combined two requirements for the applicant in one block.

Semantic intersection of clusters is also possible. Thus, the clusters “machinist” and “tractor driver” intersect in the phrases “excavator driver” and “tractor driver.”

It is necessary to pay attention to the model’s ability to combine synonyms, even in different languages, to consider phrases with spelling errors. For example, the concepts “responsibility,” “responsible employee,” and “obligation” were included in one class.

Based on the results of this algorithm, the authors formed the characteristics of the labor resources of the core and periphery of the digital economy, presented by employers when describing vacancies. The authors sorted out the lists of requirements in descending order of importance (according to the TF-IDF measure) (Saydakova, 2024).

Thus, the authors formed a profile of the applicant required by digital periphery enterprises for each enlarged category of economic activities. The main characteristics of the profiles are shown in Table 1.

Table 1 Profiles of employees in demand at digital edge enterprises

Enlarged category	Required employee profile
Water supply, drainage, and waste disposal	<ol style="list-style-type: none"> 1. Higher engineering and technical education; 2. Availability of knowledge, skills, and experience as a design engineer; 3. Direct work
Higher education	<ol style="list-style-type: none"> 1. Higher, secondary specialized education, or additional professional education; 2. Having permission to work with people; 3. Direct teaching activities
Hotels and catering	<ol style="list-style-type: none"> 1. Communication skills, good speech, and culture of communication; 2. Responsibility and punctuality; 3. Having permission to work with people; 4. Interaction with clients and stress resistance
Public administration, social security	<ol style="list-style-type: none"> 1. Secondary education, preferably higher legal or medical education; 2. Responsibility, good speech, and basic skills in using a PC and office equipment; 3. Suitability for military service
Mining	<ol style="list-style-type: none"> 1. Secondary vocational education; 2. Responsibility; 3. Knowledge of work technology; 4. Direct execution of mining operations
Health and social service provision	<ol style="list-style-type: none"> 1. Higher or secondary specialized medical education; 2. Responsibility and good speech; 3. Skills in using a PC and office equipment; 4. Direct medical practice
Information and communication	<ol style="list-style-type: none"> 1. No educational requirements; 2. Communication skills, stress resistance, responsibility; basic skills in using a PC and office equipment; 3. Installation of equipment, customer service, and work in call centers
Culture and sports	<ol style="list-style-type: none"> 1. No educational requirements; 2. Responsibility, punctuality, and communication skills; 3. Working directly with people or working with children
Energy supply	<ol style="list-style-type: none"> 1. Higher engineering and technical education; 2. Availability of knowledge, skills, and experience as an engineer; 3. Direct work on installation of electrical equipment
Manufacturing industries	<ol style="list-style-type: none"> 1. No priority educational requirements; 2. Diligence, responsibility, punctuality, and discipline; 3. The specifics of the activity are not a priority
Real estate transactions	<ol style="list-style-type: none"> 1. No priority educational requirements; 2. Specific knowledge or experience in real estate; 3. Work in sales and direct work with clients
Wholesale and retail trade	<ol style="list-style-type: none"> 1. No priority educational requirements; 2. Punctuality, responsibility, discipline, and accuracy; 3. No priority requirements for the type of work
Information technology industry	<ol style="list-style-type: none"> 1. Experience in the specialty; 2. Diligence, responsibility, punctuality, and ability to learn; 3. Carrying out work on setting up equipment, working with IC, work to support automated production
Professional, scientific, and technical activities	<ol style="list-style-type: none"> 1. Higher education; 2. Experience in the specialty; 3. Communication skills, responsibility, punctuality, and stress resistance; 4. Ability to interact with people
Agriculture	<ol style="list-style-type: none"> 1. Higher education; 2. Experience in agriculture; 3. Attentiveness, communication skills; 4. Direct execution of work
Construction	<ol style="list-style-type: none"> 1. No priority educational requirements; 2. Efficiency, communication skills, responsibility; 3. Direct construction and installation work

(continued)

Table 1 (continued)

Enlarged category	Required employee profile
Transportation and storage	<ol style="list-style-type: none"> 1. No priority educational requirements; 2. Knowledge of the car; 3. Stress resistance, communication skills, and responsibility; 4. Work with transport, work on maintenance of means of transport
Financial sector	<ol style="list-style-type: none"> 1. Higher or incomplete higher education in economics; 2. Experience in the banking sector; 3. Culture of communication, attentiveness, and responsibility; 4. Skills in searching for information in open sources; knowledge of Microsoft Word and Excel; 5. Consulting clients

Source Compiled by the authors

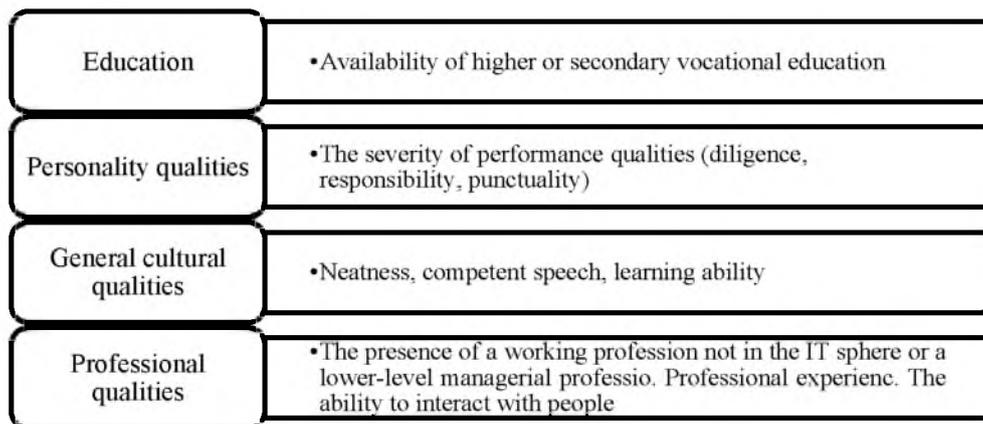


Fig. 1 Factors contributing to the decision to work in the periphery of the digital economy. Source Compiled by the authors

Profiles of the required employee are individual for each type of economic activity due to the specific features of enterprise activities. Thus, suitability for military service is necessary for public administration and social security. Medical practice is important for healthcare due to the specific nature of the activities of medical institutions.

Based on the data presented in Table 1, the authors identified factors that influence the decision-making process about working in the periphery of the digital economy (Fig. 1).

Thus, the authors justified the hypothesis of the study using the created machine learning model: for each enlarged category, the corresponding characteristics of labor resources are selected by the type of economic activity of the enterprise (employer) that operates in the periphery of the digital economy. Based on the research results, the authors identified general factors contributing to the decision to work in the field of the periphery of the digital economy. A debatable issue within the framework of this research is the formation of a profile of an employee in the periphery of the digital economy according to enlarged categories of economic activity within the territorial context.

4 Conclusion

The authors proposed a model to characterize labor resources in the periphery of the digital economy, which complements the existing methodological apparatus for studying the digital economy and labor market development. The results of using the model made it possible to create profiles of employees in the periphery of the digital economy. The results can be used to assess the level of penetration of digital technologies into the activities of enterprises.

Acknowledgements The research was carried out at the expense of a grant from the Russian Science Foundation, project No. 23-28-00640, <https://rscf.ru/project/23-28-00640/>.

Data Availability The graph of the dependence of the proportion of the explained variance on the number of components considered when using the principal component method is available at <https://figshare.com/> with an identifier <https://doi.org/https://doi.org/10.6084/m9.figshare.26075563>.

References

- Ashmarina, U. V., Gornostaeva, Z. V., Sozinova, A. A., & Savelyeva, N. K. (2023). Responsible HRM in the quality management system at enterprises of Industry 4.0 in Russia. *Proceedings on Engineering Sciences*, 5(S2), 221–234. <https://doi.org/10.24874/PES.SI.02.004>
- Belousov, Y. V., & Timofeeva, O. I. (2019). Methodology for defining the digital economy. *The World of New Economy*, 13(4), 79–89. <https://doi.org/10.26794/2220-6469-2019-13-4-79-89>
- Bogomolova, Y. I. (2012). Segmentation of contemporary labour market: Theoretical and methodological approach. *Theory and Practice of Social Development*, 8, 341–345.
- Chukov, R. S., & Zolotov, S. A. (2021). Unconditional basic income as a motivator for workers in the periphery of digital economy. *Vestnik of Lobachevsky State University of Nizhni Novgorod. Series: Social Sciences*, 4(64), 112–119. https://doi.org/10.52452/18115942_2021_4_112
- HeadHunter. (n.d.). *Official website*. Retrieved from <https://hh.ru/>. Accessed 17 March 2023
- Kaibicheva, C. I., Shurova, I. I., & Kaibichev, I. A. (2020). Center-periphery polarization of Russian Space with development of digital economy. In D. B. Solovov, V. V. Savaley, A. T. Bekker, & V. I. Petukhov (Eds.), *Proceeding of the international science and technology conference "FarEastCon 2019"* (pp. 965–973). Springer. https://doi.org/10.1007/978-981-15-2244-4_92
- Kazantseva, S. Yu., Trotsenko, V. V., Zhilyakova, M. A., & Khrustalev, V. N. (2020). Specific features of development and segmentation of the labor market as a component of economic security of the Russian Federation. *Economics and Business: Theory and Practice*, 3-1, 77–81. <https://doi.org/10.24411/2411-0450-2020-10172>
- Lin, Q., & Zhou, P. (2024). Synergy of digital economy and green economy in sustainable development policy. *Journal of Autonomous Intelligence*, 7(3). <https://doi.org/10.32629/jai.v7i3.1299>
- Lumpuva, M. A. (2018). Problems and prospects of digitalization of the Russian economy. World-system analysis. *Bulletin of the Scientific Association of Students and Graduate Students of the Faculty of History of the Perm State University for the Humanities and Pedagogy. Series: Studis Historica Juvenum*, 1(14), 127–137.
- Radaev, V. V. (2008). *Economic sociology: Handbook for higher educational institutions*. HSE University Publishing House.
- Saydakova, V. (2024). Characteristics of the workforce of the core and periphery of the digital economy declared by employers when describing vacancies. *Figshare*. <https://doi.org/10.6084/m9.figshare.25007084>
- Saydakova, V., Chuprakov, D., Sozinova, A., & Savelieva, N. (2024). Dependence of the share of the explained variance on the number of components taken into account when applying the principal components method. *Figshare*. <https://doi.org/10.6084/m9.figshare.26075563.v1>
- Shaengchart, Y., & KraiwJanit, T. (2022). Factors affecting digital economy education. *International Journal of Trade and Commerce-IIARTC*, 11(1), 16–21.
- Siqueira, E., & Fontenelle, I. (2023). Bringing labor back: Financialized inclusion and survival struggles in the periphery of the periphery. *Environment and Planning A: Economy and Space*, 56(3), 699–716. <https://doi.org/10.1177/0308518X231214945>
- Sozinova, A. A., Savelyeva, N. K., Chuprakov, D. V., & Saidakova, V. A. (2023). Personnel training for the green economy with the help of ESG management of universities in Russia and Central Asia. In E. G. Popkova, & B. S. Sergi (Eds.), *ESG management of the development of the green economy in Central Asia* (pp. 275–283). Springer. https://doi.org/10.1007/978-3-031-46525-3_30
- Sozinova, A. A., Savelyeva, N. K., Tolmachev, A. V., & Shchukina, N. V. (2023). Quality management at enterprises of agro-industrial complex (AIC) 4.0 as the basis of their digital competitiveness. *Proceedings on Engineering Sciences*, 5(S2), 235–248. <https://doi.org/10.24874/PES.SI.02.005>
- Sozinova, A. A., Glushko, O. A., Kurilova, A. A., & Menshchikova, V. I. (2023). Improvement of product quality in the AI economy: Human knowledge vs. digital technologies. *International Journal for Quality Research*, 17(4), 1253–1264. <https://doi.org/10.24874/IJQR17.04-19>
- Sozinova, A. A., Savelyeva, N. K., Chuprakov, D. V., & Saidakova, V. A. (2023). Quality management in Industry 4.0 through the movement of labor resources towards the “core” of the digital economy. *Proceedings on Engineering Sciences*, 5(S2), 341–354. <https://doi.org/10.24874/PES.SI.02.012>
- Speshilova, N. V., Gorelikova-Kitaeva, O. G., Rakhmatullin, R. R., & Rogachev, A. A. (2023). Labor resources and their role in the development of the digital economy. *Economic Sciences*, 5(222), 330–338. <https://doi.org/10.14451/1.222.330>
- Tagarov, B. J. (2020). The impact of the economy digitalization on the competitive environment: Threats to national manufacturer. *Journal of Economics, Entrepreneurship and Law*, 10(11), 2627–2636. <https://doi.org/10.18334/epp.10.11.111160>
- Tolmachev, A. V., Sozinova, A. A., Savelyeva, N. K., & Saidakova, V. A. (2023). The attraction of digital personnel and corporate training in quality management in industry 4.0 with a balance between sustainability and competitiveness. *Proceedings on Engineering Sciences*, 5(S2), 311–326. <https://doi.org/10.24874/PES.SI.02.010>
- Wallerstein, I. M. (2006). *World-system analysis: An introduction* (N. Tyukina, Transl.). Publishing House “Territory of the Future” (Series “Alexander Pogorelsky’s University Library”).
- Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., et al. (2022). Text embeddings by weakly-supervised contrastive pre-training. *arXiv*. <https://doi.org/10.48550/arXiv.2212.03533>
- Wang, L. (2022). Multilingual-E5-Large. *Hugging Face*. Retrieved from <https://huggingface.co/intfloat/multilingual-e5-large>. Accessed 9 January 2024
- Zhang, X., Ma, X., Shi, P., & Lin, J. (2021). Mr. TyDI: A multilingual benchmark for dense retrieval. *arXiv*. <https://doi.org/10.48550/arXiv.2108.08787>
- Zolotov, S. A. (2022). Digital economy: The core and the periphery. *Semiotic Studies*, 2(2), 98–103. <https://doi.org/10.18287/2782-2966-2022-2-2-98-103>