ARTICLE

TOWARDS ALGORITHM-ASSISTED CAREER MANAGEMENT – A CHALLENGE FOR NEW IMMIGRATION COUNTRIES. PREDICTING MIGRANTS' WORK TRAJECTORY USING ENSEMBLE LEARNING

Maj, J., Ruszczak, B., Kubiciel-Lodzińska, S.

Jolanta Maj / Wrocław University of Science and Technology, Faculty of Management, Department of Management Systems and Organizational Development, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland. Email: jolanta.maj@pwr.edu.pl (corresponding author)

Bogdan Ruszczak / Opole University of Technology, Faculty of Electrical Engineering Automatic Control and Informatics, Department of Informatics, Prószkowska 76 Street, 45-758 Opole, Poland. Email: b.ruszczak@po.edu.pl

Sabina Kubiciel-Lodzińska / Opole University of Technology, Faculty of Economics and Management, Department of Regional Policy and Labour Market, Prószkowska 76 Street, 45-036 Opole, Poland. Email: s.kubiciel-lodzinska@po.edu.pl

Abstract

Migration processes have emerged as crucial social, political and economic concerns, affecting societies, industries and organisations. The challenge lies in effectively utilizing immigrants' resources. This research aims to determine how AI tools can support matching migrants' skills with labour markets in host countries. We propose the application of an ensemble learning methodology. To validate this approach, we collect data to assess the career trajectories of 248 tertiary-educated Ukrainian immigrants in Poland, a new immigration destination. Various machine learning models are evaluated using the decision tree algorithm on these feature sets. To ensure credible results, a 10-fold cross-validation procedure is employed for each training process of every submodel. This research introduces an original ensemble machine learning classifier that combines pre-selected models with the highest performance, thereby reducing the number of parameters to be investigated. Its application in determining the career paths of highly skilled migrants, specifically Ukrainians, is novel. The study offers significant implications for Central Europe, notably Poland, where migration patterns and the integration of highly skilled migrants, mainly from Ukraine, are increasingly important.

Implications for Central European audience: The ensemble machine learning classifier developed in this study could aid in optimising the career paths of these migrants, combating brain waste and facilitating their successful integration into the labour market. Integrating tools like these into decision-making processes may enhance career management and contribute to Central Europe's social and economic growth.

Keywords: Career management; migration; immigrants; machine learning; ensemble learning; decision trees; labour market

JEL Classification: F22, J61, O15

Introduction

There is no doubt that artificial intelligence (AI) and machine learning (ML) are entering organisations nowadays. AI, which can be defined as a simulation of human intelligence in machines that are programmed to think and learn, has been evolving for more than 50 years. Many researchers are trying to develop proficient techniques for creating intelligent programs; the last decade has been a boom for ML, a subfield of AI focused on automating the data-driven development process. Among many other methods based on machine learning, ensemble learning (EL) is worth mentioning. It allows the delivery of well-performing models from candidates that initially appear weak. However, as has been reported for many fields, for instance, precision agriculture (Ruszczak et al., 2020) or marketing research (Ruszczak et al., 2023), it supports pattern recognition, finding and confirmation (Nalepa et al., 2021). The ensemble machine learning techniques could also significantly aid the reasoning and help find new internal data relations for labour studies. It is widely believed that redesigning the workplace to integrate AI will become necessary to stay competitive or survive (Konovalova et al., 2022). AI was first introduced in the 1950s and has experienced ups and downs. However, with the rapid development of ML technologies, it has been revitalised (Duan et al., 2019). Within organisations, AI has improved the ability to use data and make predictions and has therefore improved the decision-making process (Agrawal et al., 2018). Using AI in the decision-making process has been one of the most crucial applications in AI history, leading to the development of algorithms and algorithmic management (Lee et al., 2015). Algorithm-aided decisions have already been used for recruitment and selection, learning and development, performance analysis and management, compensation and reward and employee turnover (Yahia et al., 2021). However, the application of AI and algorithmic management is mainly limited to big organisations, and the literature on it is relatively new and not very extensive, yet rapidly gaining attention (Pan et al., 2022).

In the present paper, the EL method is described as an adjunctive decision-making method in the assessment of career trajectories of highly skilled economic migrants. The research aims to establish to what extent AI tools can support matching migrants' skills with labour markets in host countries. The analysis aims to identify the factors facilitating migrants' pursuit of skilled employment. Two critical conditions are set for the investigated method: the employed modelling architecture must strongly support model interpretability and the component models must be manageable. Thus, some candidates are evaluated and combined with the resulting classifier of the decision trees working on handcrafted sets of features, and a couple of different strategies are used to get the final model response. For the best-performing weighted voting of the candidate models, the F1 score of 0.950 is obtained and the correct classification is used for 96% of respondents using the evaluation set.

Highly skilled migrants generally come from countries with lower economic development than the host country and are often treated as “second-class” workers due to their origin (Binggeli et al., 2013; Dietz et al., 2015). One of Europe's challenges is using their skills and not letting their talents go to waste. In 2020, nearly one-third of migrants born outside the EU had a tertiary education (Eurostat, 2021). The study focused on highly skilled migrants, chosen due to their underutilised expertise in destination countries, leading to a depreciation of human capital. Employers often find it more convenient to depend solely on the immediately applicable skills of migrant workers (Crowley-Henry & Al Ariss, 2018). Conversely, harnessing the full potential of human capital and providing immigrants with opportunities to apply their skills poses a significant challenge. Given the growing importance of migratory processes in Europe, developing a tool capable of harnessing their talents is crucial for companies' socio-economic development and career management.

The study contributes to the body of knowledge on the use of ensemble learning in determining the career paths of highly skilled migrants, especially in new immigration countries. To our knowledge, such an approach has not yet been used for this purpose. Therefore, it contributes to the literature on algorithmic management, career management and human resource management (HRM). The study also contributes to migration research and HRM by using machine learning to support the job-qualification matching process.

The rest of the paper is structured as follows. In Section 1, the theoretical background of algorithmic management and algorithmic decision-making is presented and then described in more detail in the context of career management. Then, the research design is introduced in Section 2, followed by the findings in Section 3. Next, Section 4 discusses the results considering the current state of the art and presents the research implications on the ongoing debate. The paper ends with a conclusion.

1 Theoretical Background

1.1 Voluntary and forced migration to Poland

In European Union (EU) countries, migrant employment is predominantly observed in elementary occupations. This phenomenon is attributed, firstly, to the prevailing structure of labour demand for immigrants within the host countries, wherein job vacancies are concentrated in sectors that involve more straightforward tasks and are less appealing to native workers. Secondly, utilizing the qualifications of migrant workers typically necessitates an investment of time and resources, both on the part of the immigrant and the host country administration (Fernández-Reino & Rienzo, 2022). This is due, among other things, to language barriers (Liu et al., 2015), employers' attitudes towards immigrants (Tharenou & Kulik, 2020) and legal conditions (Axelsson, 2017). In contrast, employing migrant workers in elementary occupations, characterised by unskilled services with low prestige, circumvents the need for such investments and enables immediate employment upon arrival. One of Europe's challenges is using their skills and not letting their talents go to waste.

In Poland, migrants have become an essential part of the labour force. By February 2022, an estimated 2 million foreigners were in the country, of which the majority, 1.35 million according to estimates, were people from Ukraine – mainly economically active men (Duszczyk & Kaczmarczyk, 2022). This influx was primarily of a labour-intensive nature. The presence of immigrants, mainly Ukrainian nationals, on the Polish labour market is primarily due to the deficit of skilled and unskilled manual workers and it was mainly such persons that employers were looking for (Górny et al., 2013). In Poland, economic immigrants are treated mainly as additional "hands to work", a workforce in short supply on the Polish labour market. At the same time, the intellectual potential they bring with them is marginalised. Research shows that among labour migrants from Ukraine, there is also a group of people with tertiary education and their share varies from a dozen to several tens of per cent depending on the study (Górny et al., 2013). The share of people with higher education among refugees from Ukraine is even higher, as according to research, about half of them have a university degree (Pędziwiatr et al., 2022; Kubiciel-Lodzińska et al., 2024). In contrast, despite having a higher education, most refugees hold simple jobs (Duszczyk et al., 2023).

A similar situation can be observed in the Czech Republic, where immigrants were seen as a labour force to fill short-term gaps in the job market. Apart from the Blue Card (an instrument of the European Union), there is a lack of national legal instruments for the long-term use of the qualifications of highly educated people (Čada & Hoření, 2021). Immigrants are mainly employed in low-skilled jobs, regardless of their educational level. In Slovakia, on the other hand, Ukrainian citizens play vital roles across various sectors, serving as specialists, technicians, professionals, administrative staff and skilled workers in agriculture, forestry and fishing, contributing significantly to these fields (Koroutchev & Novotny, 2020). However, after 2015, one can see an increase in the share of the proportion of low-skilled workers at the expense of high-skilled workers and specialists. It is in line with the shortage of mostly cheap labour in the expanding industrial production in Slovakia.

1.2 Career management of immigrant employees

Research into career management of foreign workers has focused chiefly on expatriates sent by their employers to another country for temporary work assignments. However, immigrants, whose mobility is often long-term and involves not only career changes but also life transitions, have received little research attention (Zikic et al., 2010).

When analysing their careers, research has focused on so-called objective and subjective careers. Subjective careers represent individual experiences: attitudes, motivation and resilience (Palic et al., 2023). The objective careers represent the perception of the host society and focus on institutional and professional barriers to re-establishing their careers (Guo et al., 2021). Despite the war for talent and immigrants being a potential source of competitive advantage and strategic value (Zikic, 2015), when moving to another country, they often face exclusion from organisational talent pools (Crowley-Henry et al., 2018). Their skills, experience and qualifications are often undervalued in the host society (Pearson et al., 2012), leading them to take up employment below their qualifications and resulting in skill discounting (Treuren et al., 2021) and brain waste. However, some immigrants can find a job similar to the one they performed in their home country, working according to their qualifications (Ramboarison-Lalao et al., 2012). As O'Connor and Crowley-Henry (2020) pointed out, there is still much to learn about career actions and outcomes at the individual and contextual levels.

Therefore, the paper asks the following research question:

What information can help determine the immigrant's ability to navigate the host country's labour market and help them and organisations in career-related decision-making processes?

1.3 Using algorithms in career management

Algorithms have been implemented in the workplace, e.g., to enhance the automation of physical or cognitive tasks (Wang & Siau, 2019) and to help in the decision-making process (Lindebaum et al., 2020). So-called intelligent systems have also been implemented in

https://doi.org/10.18267/j.cebr.365

organisations to automate tasks, which until then were conducted by managers (Huang, 2022). The term “algorithmic management“ was first used by Lee et al. (2015) and it includes “data-driven decision-making made autonomously by an algorithmic system as well as human decision-making assisted by algorithmic systems” (Gagné et al., 2022, p. 4). In the following paper, the focus was put on the latter definition, where human decisions are assisted by algorithmic systems. Several studies have implied that algorithmic management often reproduces and exacerbates inequalities and injustices rather than alleviating them (Benjamin, 2019). Scholars have already touched upon topics such as algorithmic discrimination (Lara, 2022). Therefore, it is believed that until those issues are resolved, the human algorithm decision-making process will be more suitable, especially when analysing vulnerable populations such as immigrants (Hakak & Al Ariss, 2013).

As mentioned by Kinowska and Sienkiewicz (2023), management areas that were previously the sole responsibility of managers are increasingly affected by algorithmic decision-making. Mallafi and Widyantoro (2016) analysed the possibility of using prediction modelling on career management. However, their analysis focused on the selection process for employees who have the potential to become managers and did not include a specific group such as immigrants. Furthermore, certain HRM areas have received much more attention concerning the possibility of applying algorithmic management than others (Cheng & Hackett, 2021). Career management is among the latter.

This paper aims to establish whether it is possible to determine career paths using AI on an example of a unique population. A question that often arises in career management is how to use historical data. AI differs from previously used data-supported decisions (human resource information systems) as they rely primarily on descriptive analytics. Current algorithms, however, allow the use of models and data mining techniques to predict future performance, turnover and competency gaps (De Vos et al., 2009); the paper also shows career paths.

In the present paper, career management is defined as a set of activities undertaken by organisations to plan and manage the careers of their employees (Berbyuk Lindström, 2018) or individuals. We do not wish to limit ourselves to employees, especially when, for immigrants, it may be an external organisation (public administration office or job agency) that influences and manages their career paths.

2 Research Design and Methods

2.1 Data characteristics

We used purposive sampling in designing the survey. The criteria for the selection of respondents (inclusion criteria) were defined in detail: migrants had to come from the same national group (Ukraine), have the same level of education (tertiary education) and have worked in Poland for at least six months before the survey. The research was comparative. Thus, migrants from Ukraine working in Poland were recruited according to their qualifications and working below their qualifications. The quantitative survey was conducted from December 2021 to March 2022 among 248 higher-educated Ukrainians. A complete list of survey questions, choices and their encodings can be found in the Appendix. The sample characteristics are presented in Table 1.

Table 1 | Sample characteristics

Feature	Option	Number of samples	Ratio [%]
Gender	Women	100	40.3
	Men	148	59.7
Age	Up to 35 years	113	45.6
	36–44	88	35.5
	45–55	39	15.7
	Over 56 years	8	3.2
Country of graduation	Ukraine	203	81.9
	Poland	37	14.9
	Other countries	8	3.2
Year of arrival in Poland	2003 and before	12	4.8
	2004–2010	19	7.7
	2011–2015	40	16.1
	2016 and later	177	71.4
Nature of work in Poland	Work according to qualifications	101	40.7
	Work below qualification	147	59.3
	Have not worked in Ukraine	36	14.5
Length of work experience in Ukraine	1–5 years	84	33.9
	6–10 years	57	23.0
	Over 10 years	71	28.6

Source: Own elaboration

2.2 Modelling migrants' career characteristics using ensemble learning

We used the responses to the survey questions to prepare the datasets needed for the training and inference process for the individual models.

For the case analysed, the resulting model, using the given set of features, is a classifier that provides information as to whether the assessed data entry reflects an immigrant working according to qualifications (denoted as ACC, short for accordingly) or an immigrant working below qualifications (denoted as NOT).

Two research hypotheses were formulated:

H1: The field of study is crucial when securing employment aligned with qualifications in an immigration country.

H2: Work experience in Ukraine does not necessarily lead to securing a job compatible with qualifications in Poland.

It was essential to code the responses to the questions appropriately, maintain the order of the responses on the scales and then rescale the resulting variables using a standard scaler. We used the individual questions to prepare the features and sets of such features were fed into the training process for the individual models. The features were selected considering the researchers' experience of the problem being modelled. This approach was intended to enable subsequent interpretation of the prediction results: achieving a high score in the validation of a given model made it possible to treat the interpretations of a given model as reliable, but it also introduced a complication: each candidate model was based on a different set of features.

The analysis determined six candidate models based on data from the following feature sets (see the Appendix for a complete list of feature names, values and encodings):

- S1: {7i, 7j, 9a, 9b, 9c, 10a, 10b, 10c, 10d, 10e, 13, 14, 15a, 16, 17b, 17c, 18b, 20},
- S2: {7b, 7c, 7d, 7g},
- S3: {7d, 7e},
- S4: {7b, 7c, 7d, 7e, 7g, 7i, 7j, 9a, 9b, 9c, 10a, 10b, 10c, 10d, 10e, 13, 14, 15a, 16, 17b, 17c, 18b, 20},
- S5: {3a, 3b, 3c, 3d, 3e, 3f},
- S6: {5a, 5b, 5c, 5d, 5e}.

As can be seen, some feature sets overlap (S1 and S4), while others (S5 and S6) are based on entirely different features. The aim was to capture different potential underlying patterns in the data, so we used different combinations of features.

We used a decision tree algorithm to build individual models to keep the interpretability of the models high. For each decision tree generated, three hyperparameters were selected in each case: "max_depth" from the range {1:7}, "min_samples_split" from the range {0.01:1.0} and "min_samples_leaf" from the range {1,3,5,7,10,20}. For the pool of models to select the final score, we selected the candidate model with the highest F1 score for each parameter set. Information on the survey questions on which this dataset is based can be found in the Appendix. The six candidate models were trained using the decision tree algorithm and we performed a 10-fold cross-validation procedure. To verify the behaviour of the models, their outcomes were measured. For the predicted values of two classes, we included TP (true positives, when somebody works according to their qualifications and models predict the same), TN (true negatives, when models predict the same class as the real data are) and situations where the predictions failed: FP for false positives and FN for false negatives.

The performance of each model was verified using conventional metrics such as accuracy, precision and recall. Although they are popular indicators for classifiers, their interpretation concerning unevenly populated classes should be cautious. Therefore, we decided to check additional metrics: F1 score and r_{φ}.

The F1 score (which aggregates the sum of true positives, false negatives and false positives) was determined by comparing the model estimates with the results read from the reference set, according to the formula:

$$F_1 score = \frac{2TP}{2TP+FN+FP} \tag{1}$$

The second metric, r_φ, also known as the Matthews correlation coefficient, denotes

$$r_\varphi = \frac{TN \cdot TP - FP \cdot FN}{\sqrt{(TN+FN)(TP+FP)(TN+FP)(TP+FN)}} \tag{2}$$

The r_φ provides non-overly optimistic, non-inflated results and allows the evaluation of models built on unbalanced datasets (Chicco & Jurman, 2020). The r_φ is also more robust than accuracy and F1 score and summarises model performance in a single metric (Chicco et al., 2021).

The final step of our investigation was aimed at providing a single classifier. This could be one of the candidate models or a combination of their outcomes. To achieve a more accurate classifier, we decided to exploit ensemble techniques. There are two main strategies allowing the building of such an ensemble: hard and soft voting (Islam et al., 2022). The whole model development procedure was described in the Introduction.

3 Findings

3.1 Experiment with individual candidate models

We used a decision tree algorithm to build individual models in order to keep the interpretability of the models high. For each generated decision tree, three hyperparameters were selected in each case: “max_depth” from the range {1:7}, “min_samples_split” from the range {0.01:1.0} and “min_samples_leaf” from the set {1,3,5,7,10,20}. For this pool of models to evaluate the final score, the candidate model with the highest F1 score for each parameter set was selected during the cross-validation procedure.

Table 2 below summarizes the results for all six evaluated candidate models and Table 3 enumerates all the configurations of those models.

Table 2 | Results for six candidate models

Candidate model feature set	F1 score	r_{φ}	Precision	Recall	Accuracy
S1	0.891	0.803	0.919	0.886	0.900
S2	*0.837*	*0.694*	*0.858*	*0.840*	*0.847*
S3	0.690	0.388	0.726	0.689	0.709
S4	0.863	0.745	0.876	0.870	0.868
S5	*0.627*	*0.290*	*0.652*	*0.640*	*0.662*
S6	0.835	0.703	0.857	0.847	0.839

Note: Models that did not pass the acceptance threshold are shown in italics.

Source: Own elaboration

Table 3 | Basic dataset characteristics

Feature set	Parameters	Feature importance	Features of positive importance
S1	max_depth: 5, min_samples_leaf: 1, min_samples_split: 0.11	[0, 0, 0.026, 0, 0.087, 0, 0.011, 0.211, 0, 0, 0.005, 0.001, 0, 0, 0.017, 0, 0, 0]	[15a, 17b, 20, 18b, 9a, 9b, 10b]
S2	max_depth: 5, min_samples_leaf: 5, min_samples_split: 0.01	[0.022, 0.246, 0.035, 0.023]	[7b, 7c, 7d, 7g]
S3	max_depth: 3, min_samples_leaf: 5, min_samples_split: 0.21	[0.071, 0.017]	[7d, 7e]
S4	max_depth: 4, min_samples_leaf: 10, min_samples_split: 0.01	[0, 0, 0, 0, 0.07, 0, 0, 0.058, 0, 0, 0, 0, 0.013, 0, 0, 0, 0, 0, 0, 0.224, 0, 0, 0]	[17b, 18b, 9c, 7c]
S5	max_depth: 2, min_samples_leaf: 1, min_samples_split: 0.41	[0.0, 0.0, 0.025, 0.0, 0.0, 0.019]	[3c, 3f]
S6	max_depth: 2, min_samples_leaf: 1, min_samples_split: 0.01	[0.161, 0.0, 0.009, 0.051, 0.0]	[5a, 5c, 5d]

Source: Own elaboration

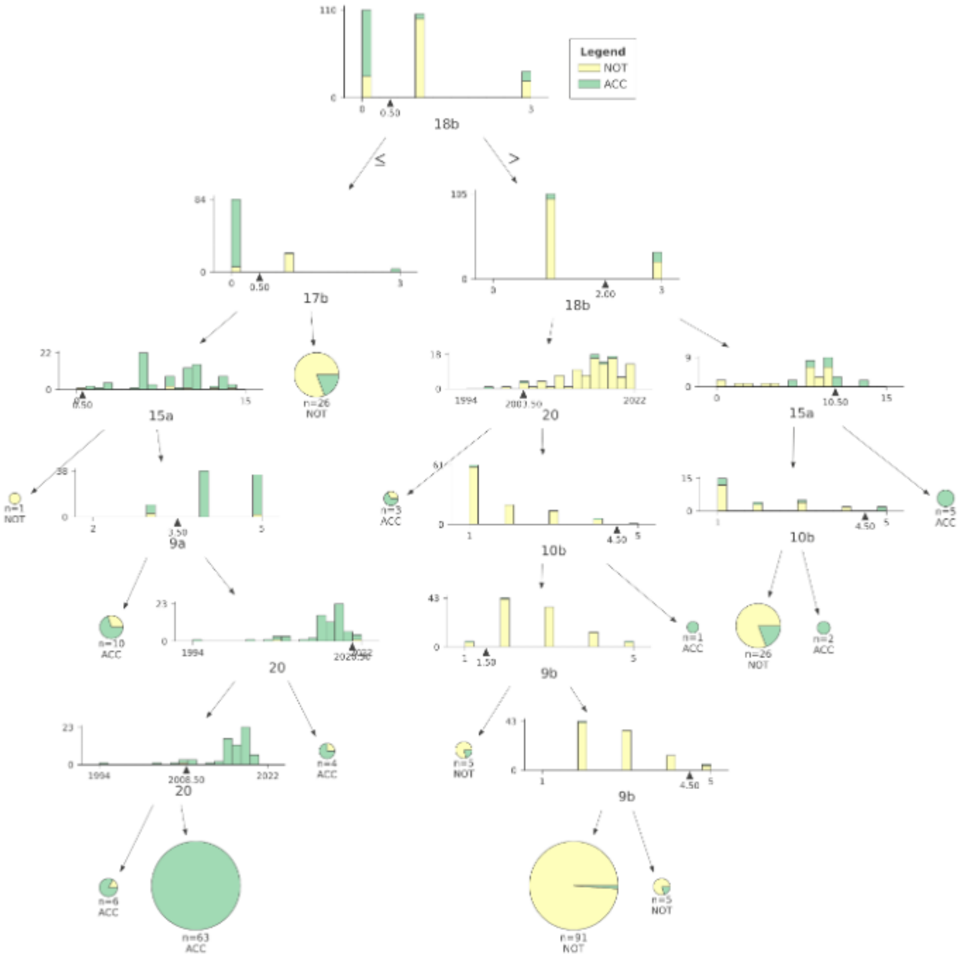
As can be seen, not all the features provided for the candidate models contribute positively to the final prediction of most of the models. For instance, the model that was trained on the set S4, which had the most features, had the best results when 4 out of the 23 were used.

Interestingly, the highest score among the candidate models, measured using cross-validation with 10 folds, was obtained for the first model (S1; F1 score = 0.891) and not for the set that had access to the largest number of learning features (S4; F1 score = 0.863).

The decision tree for the set of features fed into the algorithm is presented in Figure 1. This should aid the interpretation process by providing a detailed understanding of which questions were most important in calculating the model response. It also allows us to repeat the modelled pattern and use it with upcoming data.

In the example presented, it can also be seen that the main questions asked were 18b (change of profession between working in Ukraine and Poland), 17b (employment in accordance with qualifications in Ukraine), 15a (field of study) and 20 (length of stay in Poland). These questions for this scenario address most of the immigrants interviewed. Each of the models returns a binary response – the answer to the main question of the research: whether the respondent is working according to their competence level (ACC) or whether they are not (NOT). Only one decision tree for the best-performing model is shown (Figure 1), as the other trees are subject to a similar manner and interpretation.

Figure 1 | Visualization of best-performing candidate classifier



*. Model features that contribute to the decision are 9a, 9b, 10b, 15a, 17b, 18b, 20; and target classes are ACC (green) and NOT (yellow).

Source: Own elaboration

3.2 Ensemble learning results

The models presented above provided promising results, but for the main experiment, we decided to check whether their combined work could produce even better predictions. For this ensemble model, we used only the models that previously achieved a resulting F1 score higher than 0.75.

For the four component models, we then built the ensemble model that reflected the final prediction and tested the following two strategies to determine it:

- equally weighted voting for each of the candidate models, with a different number of votes determining the acceptance of the answer;
- weighted voting, where the F1 score read for the candidate models was used for weighting and then also determining the optimal number of positive responses required from the candidate models (scores multiplied by their F1 score and only then used to determine the final decision).

Table 4 shows a summary of the results obtained for both strategies and the different numbers of votes required for a given outcome. The highest score for both strategies is obtained for the variant where at least three candidate models were required to indicate a positive result to determine the final score. The use of weighted voting (strategy b) provided a slight additional improvement in the F1 score readout, with a final readout of 0.95 and an r_φ of 0.916, indicating 96% accuracy in identifying whether the immigrant being tested would work as qualified.

Table 4 | Results for ensemble models of two different voting strategies

Voting strategy	Number of positive votes	F1 score	r_φ	Precision	Recall	Accuracy
a)	1	0.745	0.561	0.594	*1.000*	0.722
	2	0.908	0.844	0.846	0.980	0.919
	3	*0.946*	*0.909*	0.933	0.960	*0.956*
	4	0.878	0.825	*1.000*	0.782	0.911
	5	0.577	0.537	*1.000*	0.406	0.758
b)	1	0.808	0.676	0.678	**1.000**	0.806
	2	0.934	0.888	0.892	0.980	0.944
	3	**0.950**	**0.916**	0.960	0.941	**0.960**
	4	0.768	0.704	**1.000**	0.624	0.847
	5	0.577	0.537	**1.000**	0.406	0.758

Note: For each strategy, five variants were tested for the different required number of models voting for the ACC class. The best configuration is shown in italics for the first voting strategy and in bold for the second one.

Source: Own elaboration

What may also be interesting is that it is possible to configure the ensemble differently. If high prediction precision is required, a model with at least the required four votes per class would provide 100% precision for such predictions, whereas a model based on only one submodel required to indicate those working according to qualifications would provide 100% sensitivity for their identification, but with much lower precision and accuracy for such classification.

4 Discussion

The proposed ensemble learning method enables several suboptimal machine learning models to be combined into the final classifier of better performance. Although these candidate submodels could be considered weaker, they still provide some grasp of the patterns for some parts of the investigated dataset. Two voting strategies responsible for the final model decision have been tested to find their optimal aggregation. The main advantage of this method is that the interpretive ability does not get lost and the benefits of the found patterns can be utilised. After the annotated dataset was subjected to the presented ensemble model, we verified that its characteristics could be accurately resembled. Thus, we stated that it could be applied to the final classifier for the upcoming data, and long before information about the future career of the interviewed immigrants was obtained, it could be estimated whether they would adapt to the labour market or not. A detailed analysis of the decision tree makes it possible to identify factors conducive to jobs that match the migrants' qualifications. Among them are pursuing skilled employment in the country of origin, field of study, oral proficiency in Polish and length of stay in Poland.

4.1 Theoretical contribution

The research enhances the existing body of knowledge concerning the utilisation of artificial intelligence and machine learning in discerning the career trajectories of proficient migrants, with a specific focus on individuals from Ukraine. As far as we are aware, this particular tool has not been previously employed for such a purpose. The investigation identifies variables that facilitate optimal utilisation of the skills possessed by migrants with tertiary education. The paper provides insights into the specific migrant qualifications that warrant attention in order to mitigate skill discounting and brain waste. The study contributes to the body of knowledge by demonstrating the use of artificial intelligence and machine learning techniques for determining the occupational paths of highly skilled migrants, specifically focusing on Ukrainian immigrants. This novel application of AI and machine learning tools for this purpose adds to the existing research landscape. The contribution also involves the identification of specific factors that play a role in determining the career trajectories of highly skilled migrants.

4.2 Practical contribution

At the end of 2022, there were more than 2 million migrants in Poland (Duszczyk & Kaczmarczyk, 2022), with estimations for an additional million refugees approaching (Duszczyk et al., 2023). Therefore, a great potential for an application like the one presented in this study was found. Brain waste means losses not only for the migrant, who does not have the opportunity to use his/her high skills but also for the host country, which cannot absorb the knowledge that foreigners (both economic migrants and refugees) bring with them. In the proposed model, only a part of the migrant's activity is shown, but it is crucial to prepare a tool based on machine learning that would support the integration of migrants/refugees and propose specific support.

There is an opportunity to use the method to verify and better explain findings from structured interviews. In the described study, the development of an algorithm that could indicate what career path a highly skilled migrant might take would thus facilitate, among other things, career counselling for them. It could be used to support the interviewing process carried out by a human being, who, unlike a machine, is better able to assess the aspirations and opportunities of a given migrant. Brain waste means losses not only for migrants, who do not have the opportunity to use their high skills, but also for the host country, which is not able to absorb the knowledge that foreigners (both economic migrants and refugees) bring with them. The conducted research also has important practical, managerial implications. It shows that time-consuming, labour-intensive processes such as document analysis, competency identification, etc., can be minimised so that the decision-making process can be improved through better-processed information (Binggeli et al., 2013). It also shows that using ML for managing career paths may result in higher effectiveness and efficiency through reducing costs and streamlining HRM processes and that ML-supported decision-making processes may well be used in HRM.

4.3 Ethical concerns

Artificial intelligence algorithms possess the capability to perform tasks such as visa application rejection or matching a migrant's identity with that of a suspected terrorist, lacking transparent elucidation of the decision-making process employed by the machine. This scenario can arise due to using unsupervised learning algorithms, allowing machines to autonomously learn, recognise patterns and make predictions that may deviate from human decision-making tendencies (Graves & Clancy, 2019).

4.4 Limitations of the study

The method has been tested on the presented dataset, but for its further development, it would be advisable to evaluate it with further datasets containing other types of data.

The analysis carried out has made it possible to identify questions that can be asked of migrants to determine their ability to navigate the host country's labour market and help in career-related decision-making processes. According to the model, in order to determine career opportunities, migrants should be asked about jobs compatible/incompatible with their qualifications in their country of origin, change in the nature of work as a result of their migration, their field of study, length of stay in Poland, and oral and written knowledge of the Polish language. A weakness of the research is that it was conducted only on a relatively small group of migrants (248 individuals). However, primary data are difficult to collect as it requires the implementation of direct surveys among the migrant community (Górny & Napierała, 2016). In addition, they covered only one nationality group – people originating from Ukraine. A completely different picture could be obtained by analysing the career paths of migrants from African or Asian countries, where differences in education, a greater language barrier and cultural and religious differences can significantly affect the career fate of immigrants in the host country.

On the other hand, one could say that providing a method that could reach a satisfactory performance for a relatively small dataset is valuable. Such cases are widespread, as retrieving data for large populations is usually rather difficult and costly.

4.5 Direction for further research

How does a highly skilled migrant's career path evolve? This is a challenge for future analysis to combine individual factors with macro, mezzo-economic and organizational-level factors. The conducted study used quantitative data. However, we are aware that the migrant's final career path choice is also influenced by many individual factors, such as social networks, career aspirations and plans. Social media data from Facebook and LinkedIn could also be considered. However, this is challenging, as these data are used for analyses of population movements (Goglia et al., 2022), whereas it may be difficult for a specialized issue such as the use of qualifications.

In the case of migration research, which deals with highly individual decisions such as career paths, the best data are primary data. Based on the research carried out so far, we are aware that it is possible to identify at least several pathways for highly skilled migrants and their behaviour on the host country labour market.

For effective use of the model, it is necessary to have primary data directly from migrants, e.g., level of education, field of study, work experience in the country of origin, knowledge of the host country language, knowledge of another foreign language, willingness to retain, etc. In addition, it is necessary to collate this information with the current situation on the labour market (e.g., demand for professions and qualifications).

Furthermore, it would also be interesting to analyse the political context and the migration policies pursued in the country in question, as these may also have an impact on the paths that highly skilled migrants take or can take. Also, the migrants' individual attitudes, ambitions and goals are relevant to their career choices and their aspiration to use their qualifications in the country of immigration. With the data currently available, it is not possible to predict what steps a migrant who finds themselves in a certain situation will take. Further work could also attempt to perform a similar process but for less structured data and apply large language models (LLM) to interpret respondents' open-ended answers.

Conclusion

Human activities are constantly generating significant amounts of data, but these are used to a fairly limited extent to manage migration, which has become one of the key challenges. This applies to both voluntary and involuntary migration. Big data analytics and AI technologies are seen as cutting-edge solutions to current and emerging social, economic and governance challenges, including their prediction, and can be used by politicians but also on a smaller scale by organizations and managers.

Our model confirmed H1, which is related to the field of study versus the possibility of obtaining a job in line with qualifications. With more detailed analysis, it was shown that taking up employment consistent with qualifications is facilitated by having a degree. Regarding hypothesis H2, the algorithm indicated that work experience from Ukraine is conducive to taking up employment in line with qualifications, but on the condition that the migrant also worked in line with qualifications in their country of origin.

Big data analytics is also providing social scientists with new insights into migration research. Such analysis makes it possible to study temporary or circular migration patterns and monitor in real time public opinion and media discourse on migration. The new knowledge generated from migrant data also raises challenges to ethical principles. A huge challenge for migration research is the availability of data relating to migrants. In the case of Poland, these are of quite low quality, i.e., there is a lack of data on, e.g., migrants' occupational profiles and their education. While the algorithm prefers quantitative data, qualitative data are preferred in migration studies aimed at capturing, among other things, factors influencing migration. Policy decision-making is much more complex than technical decision-making, meaning that decision-makers often compromise on critical issues and act on social perceptions or fears (e.g., this applies to the issue of accepting refugees or allowing migrants to the labour market) (Bianchi & Saab, 2019).

Data that are difficult to interpret may be analysed using tools like the one described in the paper, as it may foster the decision-making process by showing hidden meanings and patterns, which may be used in the migrant career management process. Using migrants' talents and matching their education to the needs of local labour markets is a massive challenge for host countries. On the one hand, this will make it possible to meet the needs of the job markets of developed countries with ageing populations. On the other hand, it will satisfy the professional aspirations of migrants.

Acknowledgement

Funding: This work was supported by the Opole University of Technology under Grant DELTA 2.

Conflict of interest: The authors hereby declare that this article was not submitted or published elsewhere. The authors do not have any conflict of interest.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: the simple economics of artificial intelligence*. Boston: Harvard Business Press.

Axelsson, L. (2017). Living within temporally thick borders: IT professionals' experiences of Swedish immigration policy and practice. *Journal of Ethnic and Migration Studies*, *43*(6), 974–990. https://doi.org/10.1080/1369183X.2016.1200966.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. United Kingdom: Polity Press.

Bianchi, A., & Saab, A. (2019). Fear and international law-making: An exploratory inquiry*. Leiden Journal of International Law*, *32*(3), 351–365. https://doi:10.1017/S0922156519000177.

Binggeli, S., Dietz, J., & Krings, F. (2013). Immigrants: A forgotten minority. *Industrial and Organizational Psychology*, *6*(1), 107–113. https://doi: 10.1111/iops.12019.

Berbyuk Lindström, N. (2018). Cross-Cultural Design for Employability: Mobile Support for Healthcare Professionals. In: Rau, Pl. (eds) *Cross-Cultural Design. Methods, Tools, and Users: 10th International Conference, CCD 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part I 10* (pp. 314–326). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-92141-9_24, Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 1–13. https://doi: 10.1186/s12864-019-6413-7.

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData mining*, *14*(1), 1–22. https://doi.org/10.1186/s13040-021-00244-z.

Cheng, M. M., & Hackett, R. D. (2021). A critical review of algorithms in HRM: Definition, theory, and practice. *Human Resource Management Review*, *31(*1), 100698. https://doi.org/10.1016/j.hrmr.2019.100698.

Crowley-Henry, M., & Al Ariss, A. (2018). Talent management of skilled migrants: Propositions and an agenda for future research. *The International Journal of Human Resource Management*, *29*(13), 2054-2079. https://doi.org/10.1080/09585192.2016.1262889

Crowley-Henry, M., O'Connor, E., & Al Ariss, A. (2018). Portrayal of skilled migrants' careers in business and management studies: A review of the literature and future research agenda. *European Management Review*, *15*(3), 375–394. https://doi.org/10.1111/emre.12072.

Čada, K., & Hoření, K. (2021). Governing Through Rituals: Regulatory Ritualism in Czech Migration and Integration Policy. In: Federico, V., Baglioni, S. (eds) *Migrants, Refugees and Asylum Seekers' Integration in European Labour Markets* (pp. 115–134). IMISCOE Research Series. Cham: Springer. https://doi.org/10.1007/978-3-030-67284-3_6.

De Vos, A., Dewettinck, K., & Buyens, D. (2009). The professional career on the right track: A study on the interaction between career self-management and organizational career management in explaining employee outcomes. *European Journal of Work and Organizational Psychology*, *18*(1), 55–80. https://doi.org/10.1080/13594320801966257.

Dietz, J., Joshi, C., Esses, V. M., Hamilton, L. K., & Gabarrot, F. (2015). The skill paradox: Explaining and reducing employment discrimination against skilled immigrants. *The International Journal of Human Resource Management*, *26*(10), 1318–1334. https://doi.org/10.1080/09585192.2014.990398.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

Duszczyk, M., & Kaczmarczyk, P. (2022). War and migration: the recent influx from Ukraine into Poland and possible scenarios for the future. *CMR Spotlight, 4*(39), 1–13.

Duszczyk, M., Górny, A., Kaczmarczyk, P., & Kubisiak, A. (2023). War refugees from Ukraine in Poland–one year after the Russian aggression. Socioeconomic consequences and challenges. *Regional Science Policy & Practice*, *15*(1), 181–199. https://doi.org/10.1111/rsp3.12642.

Eurostat. (2021), One third of migrants in the EU have a degree. Retrieved: December 10, 2023, from: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210707-2.

Fernández-Reino, M., & Rienzo, C. (2022). Migrants in the UK labour market: An overview. Retrieved December 10, 2023, from https://migrationobservatory.ox.ac.uk/resources/briefings/migrants-in-the-uk-labour-market-an-overview/

Gagné, M., Parent-Rocheleau, X., Bujold, A., Gaudet, M. C., & Lirio, P. (2022). How algorithmic management influences worker motivation: A self-determination theory perspective. *Canadian Psychology/Psychologie Canadienne*, *63*(2), 247–260. https://doi.org/10.1037/cap0000324.

Goglia, D., Pollacci, L., & Sîrbu, A. (2022). Use of non-traditional data sources to nowcast migration trends through Artificial Intelligence technologies. In *Measuring Migration Conference 2022 Conference Proceedings* (pp. 15). Transnational Press London.

Górny, A., Kaczmarczyk, P., Napierała, J., & Toruńczyk-Ruiz, S. (2013). *Raport z badania imigrantów w Polsce [Report from immigrant study in Poland*]. Warsaw: Ośrodek Badań nad Migracjami Fundacja.

Górny, A., & Napierała, J. (2016). Comparing the effectiveness of respondent-driven sampling and quota sampling in migration research. *International Journal of Social Research Methodology*, *19*(6), 645–661. https://doi.org/10.1080/13645579.2015.1077614.

Graves, A., & Clancy, K. (2019). Unsupervised learning: The curious pupil. *DeepMind blog*, 25.

Guo, G. C., Hakak, L. T., & Al Ariss, A. (2021). Institutional logics and foreign national origin based inequality: The case of international migrant employees. *Human Resource Management Review*, *31*(1), 100706. https://doi.org/10.1016/j.hrmr.2019.100706.

Hakak, L. T., & Al Ariss, A. (2013). Vulnerable work and international migrants: A relational human resource management perspective. *The International Journal of Human Resource Management*, *24*(22), 4116–4131. https://doi.org/10.1080/09585192.2013.845427.

Huang, H. (2022). Algorithmic management in food-delivery platform economy in China. *New Technology, Work and Employment*, *38*, 185–205. https://doi.org/10.1111/ntwe.12228.

Islam, M. M., Usman, M. U., Newaz, A., & Faruque, M. O. (2022). Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements. *IET Smart Grid, 6*(3), 219–232. https://doi.org/10.1049/stg2.12091.

Kinowska, H., & Sienkiewicz, L. J., (2023) Influence of algorithmic management practices on workplace well-being – evidence from European organisations. *Information Technology & People*, *36*(6), 21–42. https://doi.org/10.1108/ITP-02-2022-0079.

Konovalova, V., Mitrofanova, E., Mitrofanova, A., & Gevorgyan, R. (2022). The Impact of Artificial Intelligence on Human Resources Management Strategy: Opportunities for the Humanisation and Risks. *WISDOM*, *2*(1), 88–96.

Koroutchev, R., & Novotny, L. (2020). International migration to an economically lagging EU region: case study of Ukraine and Eastern Slovakia. *Geographia Cassoviensis*, *XIV*(2), 144–163.

Kubiciel-Lodzińska, S., Golebiowska, K., Pachocka, M., & Dąbrowska, A. (2024). Comparing pre-war and forced Ukrainian migrants in Poland: Challenges for the labour market and prospects for integration. *International Migration, 62*, 236–251. https://doi.org/10.1111/imig.13213.

Lara, C. S. (2022). Gestión algorítmica empresarial y tutela colectiva de los derechos laborales [Algorithmic management and collective protection of labour rights]. *Cuadernos de Relaciones Laborales, 40*(2), 283–300. https://doi.org/10.5209/crla.79417.

Lee, Y., Lee, J., & Hwang, Y. (2015). Relating motivation to information and communication technology acceptance: Self-determination theory perspective. *Computers in Human Behavior*, *51,* 418–428. https://doi.org/10.1016/j.chb.2015.05.021.

Lindebaum, D., Vesa, M., & Den Hond, F. (2020). Insights from "the machine stops" to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Academy of Management Review*, *45*(1), 247–263. https://doi.org/10.5465/amr.2018.0181.

Liu, X., Gao, L., Lu, J., & Wei, Y. (2015). The role of highly skilled migrants in the process of inter-firm knowledge transfer across borders. *Journal of World Business, 50*(1), 56–68. https://doi.org/10.1016/j.jwb.2014.01.006.

Mallafi, H., & Widyantoro, D. H. (2016). Prediction modelling in career management. In *2016 International Conference on Computational Intelligence and Cybernetics* (pp. 17–21). Makassar, Indonesia. https://doi.org/10.1109/CyberneticsCom.2016.7892560.

Nalepa, J., Myller M., Tulczyjew L., & Kawulok M. (2021). Deep Ensembles for Hyperspectral Image Data Classification and Unmixing, *Remote Sensing 13*(20), 1–29. https://doi.org/10.3390/rs13204133.

O'Connor, E. P., & Crowley-Henry, M. (2020). From home to host: The instrumental kaleidoscopic careers of skilled migrants. *Human Relations*, *73*(2), 262–287. https://doi.org/10.1177/0018726719828.

Pan, Y., Froese, F., Liu, N., Hu, Y., & Ye, M. (2022). The adoption of artificial intelligence in employee recruitment: The influence of contextual factors. *The International Journal of Human Resource Management*, *33*(6), 1125–1147. https://doi.org/10.1080/09585192.2021.1879206.

Palic, D., Nardon, L., & Hari, A. (2023). Transnational sensemaking narratives of highly skilled Canadian immigrants' career change. *Career Development International*, *28*(4), 392–405. https://doi.org/10.1108/CDI-06-2022-0182.

Pearson, J., Hammond, M., Heffernan, E., & Turner, T. (2012). Careers and talents not to be wasted: Skilled immigrants' journeys through psychological states en route to satisfying employment. *Journal of Management Development, 31*(2), 102–115. https://doi.org/10.1108/02621711211199458.

Pędziwiatr, K., Stonawski, M., & Brzozowski, J. (2022). *Imigranci Ekonomiczni i przymusowi w Krakowie w 2022 roku. [Economic and Forced Immigrants in Kraków in 2022]*. Kraków: Obserwatorium Wielokulturowości i Migracji.

Ramboarison-Lalao, L., Al Ariss, A., & Barth, I. (2012). Careers of skilled migrants: Understanding the experiences of Malagasy physicians in France. *Journal of Management Development, 31*(2), 116–129. https://doi.org/10.1108/02621711211199467.

Ruszczak, B., Smykała, K., & Dziubański, K. (2020). The detection of Alternaria solani infection on tomatoes using ensemble learning, *Journal of Ambient Intelligence and Smart Environments*, *12*(5), 407–418. https://doi.org/10.3233/AIS-200573.

Ruszczak, B., & Rudnik, K. (2023). Ensemble-based versus expert-assisted approach to carbon price features selection. *Communication Papers of the 18th Conference on Computer Science and Intelligence Systems, ACSIS* (pp. 251–256). https://doi.org/10.15439/2023F8389.

Tharenou, P., & Kulik, C. T. (2020). Skilled migrants employed in developed, mature economies: From newcomers to organizational insiders. *Journal of Management*, *46*(6), 1156–1181. https://doi.org/10.1177/0149206320921229.

Treuren, G. J., Manoharan, A., & Vishnu, V. (2021). The gendered consequences of skill-discounting for migrants. *Journal of Industrial Relations, 63*(1), 73–97. https://doi.org/10.1177/0022185620951830.

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management* (JDM), *30*(1), 61–79. https://doi.org/10.4018/JDM.2019010104.

Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. *IEEE Access*, *9,* 60447–60458. https://doi.org/10.1109/ACCESS.2021.3074559.

Zikic, J. (2015). Skilled migrants' career capital as a source of competitive advantage: Implications for strategic HRM. *The International Journal of Human Resource Management*, *26*(10), 1360–1381. https://doi.org/10.1080/09585192.2014.981199.

Zikic, J., Bonache, J., & Cerdin, J. L. (2010). Crossing national boundaries: A typology of qualified immigrants' career orientations. *Journal of Organizational Behavior*, *31*(5), 667–686. https://doi.org/10.1002/job.705.

Appendix

The answers to the following questions provided by immigrants were later encoded, rescaled (using standard scaling) and passed through the training loop. The table below presents the questions, possible answers and the sets in which the questions were later included.

No.	Question (and later model feature)	Options	Feature included in set
3a	How did you found a job? - Through a placement firm	yes, no	S5
3b	Through friends, family	yes, no	S5
3c	Through an advertisement on the internet	yes, no	S5
3d	Through an ad in the press	yes, no	S5
3e	Through an ad on social media	yes, no	S5
3f	Through other channels	yes, no	S5
5a	How did the employer (or placement agency) verify your qualifications, education/qualifications for the job? - I submitted a college diploma	yes, no	S6
5b	I submitted a document on completion of vocational course(s)	yes, no	S6
5c	The employer did not check my education/qualifications	yes, no	S6
5d	Other	yes, no	S6
5e	I don't know	yes, no	S6
7b	When working in Poland, the most important thing for me is the opportunity to get an attractive salary.	7 options on a scale	S2, S4
7c	When working in Poland, it is important for me to be able to work in accordance with my profession (education).	7 options on a scale	S2, S4
7d	I have many Polish friends in Poland and I can use their help when looking for a job.	7 options on a scale	S2, S3, S4
7e	I have many Ukrainian friends in Poland and I can use their help when looking for a job.	7 options on a scale	S3, S4
7g	Recognition of education obtained in Ukraine is very difficult in Poland.	7 options on a scale	S2, S4
7i	I intend to stay in Poland permanently.	7 options on a scale	S1, S4
7j	I intend to return permanently to Ukraine.	7 options on a scale	S1, S4
9a	Polish language skills (speaking)	5 options on a scale	S1, S4
9b	Polish language skills (writing)	5 options on a scale	S1, S4
9c	Polish language skills (reading)	5 options on a scale	S1, S4
10a	English language skills	5 options on a scale	S1, S4
10b	German language skills	5 options on a scale	S1, S4
10c	Russian language skills	5 options on a scale	S1, S4

10d	Another language	5 options on a scale	S1, S4
10e	What other languages do you know?	12 different repeating answers	S1, S4
13	Gender	M, F	S1, S4
14	Age	4 ranges	S1, S4
15a	Field of study completed		S1, S4
16	Graduation country	3 options	S1, S4
17b	Did you work in Ukraine according to your qualifications?	3 options	S1, S4
17c	Length of work experience	4 ranges	S1, S4
18b	Change of profession (between working in Ukraine and Poland)	3 options	S1, S4
20	Length of stay in Poland	23 different repeating answers	S1, S4

The research article passed the double-blind review process. | Received: 15 November 2023**; Revised:** 15 February 2024**; Accepted:** 15 February 2024**; Available online:** 7 May 2024**; Published in the regular issue:** 31 December 2024.