

# Machine learning algorithms for predicting unemployment duration in Russia

Anna A. Maigur\*

*Gaidar Institute for Economic Policy, Moscow, Russia*

---

## Abstract

Predictions of the individual unemployment duration will allow to distribute target support while searching for a job more effectively. The paper uses survival models to predict the unemployment duration based on data from Russian employment centers in 2017–2021. The dataset includes socio-demographic characteristics, such as age, gender, education level, etc., as well as the job search duration. Two models' forecasts are investigated: the proportional and the non-proportional hazards models. Both models take into account censored data, but only the second one captures nonlinear dependencies and the disproportionate influence of independent variables over time. The forecast quality is estimated with the C-index, equality of which to 1 indicates the most accurate forecast. The highest index value is demonstrated by the non-proportional hazards model (0.64). Moreover, it was found that variable that contributes the most to the prediction quality is region of a job search so that job-search time is heterogeneous among different regional labour markets. To sum up, forecast quality is quite high and stable over time and the implementation of model forecasts by employment centers will increase their efficiency.

*Keywords:* unemployment duration, survival analysis, machine learning models.

*JEL classification:* C34, C41, C45, C52, C53, J64, J68.

---

## 1. Introduction

It is crucial to understand which groups of people are the most vulnerable during their job search in order to head target support and efficiently distribute financial resources to employment centers. A model that provides valid forecast on duration of the individual unemployment could become a handy tool for them.

The problem of modelling duration is not obvious and draws some attention from statisticians. One of the characteristics of the duration data is censorship.

---

\* E-mail address: ann.maigur@mail.ru

Unemployment duration could be censored if it is not reliably known whether the unemployed found a job at the end or not. The reasons for that could be different including relocation, persistent absenteeism or some other, so these observations bring information not about the exact time of a job search but about the minimal one. They also should be included in the analysis but in a different way. Survival analysis methods are mostly used as an empirical base for this kind of research as they enable full coverage of all the data features. The most widespread model of this class of methods is the Cox proportional hazards model, the main premise of which is the proportionality of the risks for different groups of observations throughout the entire period. The model is popular because its results can be easily interpreted, but it has many limitations except for proportional hazards, the main one of which is the assumption of a linear dependence of the variables. Machine learning methods make it possible to evaluate models without such strict restrictions. Those methods usually present combination of classical machine learning models and survival analysis assumptions. For example, survival random forest is a random forest model that takes into account censored data. The other example is the non-proportional hazards model. Following Kvamme et al. (2019), hereafter it is referred to as Cox-Time model. It is a neural network where loss function is formulated as a survival analysis problem.

Classical statistical methods are usually used for structural analysis to find out the reasons for long-term unemployment. Some papers showed that unemployment duration is heterogeneous among individuals and depends on lots of factors including some personal characteristics (Kooreman and Ridder, 1983; Luboyova and van Ours, 1997; Grogan and Van Den Berg, 2001). A lot of research was aimed at investigating factors of the unemployment duration on an individual level (Jolianis et al., 2020; Stern, 1989; Luboyova and van Ours, 1997; Grogan and Van Den Berg, 2001; Denisova et al., 2003) revealing that age and education are among the most influential. Also, other authors showed that gender (Denisova et al., 2003; Bratberg and Nilsen, 2000; Grogan and Van Den Berg, 2001), some institutional factors (Bover et al., 2002) and economic conjuncture also carry great impact.

Still, very few papers were concentrated on elaborating a model to predict unemployment duration. When evaluating a predictive model, it is possible to use not only classical statistical methods, but also machine learning methods, that complicate the interpretability of the obtained results, but improve the quality of the forecast. One of the papers that presented a model for predicting a job search time is Boškoski et al. (2021). It evaluates a neural network using Bayesian methods on data from the employment centers in Slovenia to predict the distribution of a job search duration. Moreover, several papers tried to solve classification problem estimating the probability of long-term unemployment (Desiere et al., 2019). In those studies, classical machine learning models that do not consider censorship were applied. Furthermore, those models do not forecast an exact unemployment duration but a probability that this duration would exceed some predefined cut-off.

To sum up, there is a gap in studies that look into a problem of forecasting job search time based on statistical and machine learning models that consider data censorship. This gap will be partially addressed by this research. The purpose

of the paper is to elaborate a model to efficiently predict individual unemployment duration. The data is an aggregation of personnel files from employment centers all over Russia from the start of 2017 till the first half of 2020. The papers that used the same data (Nivorozhkin, 2006; Denisova et al., 2003) were focused on estimating a structural model for one or several regions. The most recent papers that conducted structural non-parametric (Giltman et al., 2022) and parametric analysis (Maigur, 2023) covered all the presented regions but left alone predictive analysis.

Using a model that forecasts the unemployment duration would allow employment centers to allocate their resources more effectively and channel targeted assistance to the most vulnerable segments of the population. Moreover, solving the regression problem rather than the classification one gives more scope for interpreting the forecast. This would potentially contribute to making more informed decisions in the field of targeted assistance in the context of various structural economic shifts.

The paper is structured as follows. Section 2 provides a literature review. Section 3 details the methodology and metrics used to compare the quality of predictions. In Section 4 used data is reported, while in Section 5 descriptive analysis is given. Section 6 gives the models' specification, and models estimation results are presented in Section 7.

## 2. Literature review

Long-term unemployment (LTU) is destructive as it has a negative impact on both the unemployed and the whole economy. Evaluating which factors affect unemployment duration has been studied a lot in recent decades. The most popular theoretical approach for describing this phenomenon is a job search model (Foley, 1997). It is assumed that the probability to find a job comprises two parts: the probability of receiving an offer and the probability of accepting it. The former depends on individual characteristics of the unemployed such as age, gender, education and others. The last depends on expected salary. The short form of the model assumes to estimate the impact of all these factors on the final probability of finding a job.

Survival analysis is used as an empirical basis for estimations. Survival analysis introduces a few specific terms such as survival function, hazard function and median survival time. All of these are used to estimate duration models. Survival function shows the probability that an individual has not found a job before a certain time. Hazard function estimates a probability that the individual would find a job in this period while not having done so before and median survival time defines the time period when the estimated survival function equals to 0.5. Survival analysis enables us to overcome the censorship problem while censored data points are common for such type of data. The information about job search is not comprehensive, and sometimes it is unclear why an individual ceased to visit an employment center. Moreover, it is not even obvious if he found a job or ended up leaving the workforce.

Many studies showed that individual characteristics, local labor market conditions and institutional arrangement are important predictors of model unemployment duration. Age (Kooreman and Ridder, 1983), gender (Denisova et al., 2003)

and education (Stern, 1989; Luboyova and Van Ours, 1997; Grogan and Van Den Berg, 2001) are listed among the most significant ones.

Nevertheless, the issue of predicting unemployment duration gained attention only recently. Just a few papers were focused on predicting unemployment duration and even fewer used machine learning models but instead favored statistical analysis. The reason is that machine learning methods are not widely used in the analysis of social policy data (Desiere et al., 2019) because they are considered to work as black-box methods. Nevertheless, predicting unemployment duration can contribute to building a strategy for targeted assistance while searching for a job.

In some papers models for predicting probability of long-term unemployment are built. Their classification problem is solved where the target variable is probability that unemployment duration will exceed some period of time (best practices are to use 6 months or 12 months as a time period). This problem formulation has grounds because long-term unemployed people are the most vulnerable, and identifying those could reduce negative consequences of unemployment to the economy (Zhao, 2020). Firstly, the unemployed have to lower their economic activity as they are deprived of some of their income. This leads to a drop in the aggregated demand, GDP and tax income as well as to an increase in expenses connected with unemployment benefits. Secondly, people who stay unemployed for a long time lose professional skills that make it even more difficult to find a job. As a result, they end up with less paid jobs than those who spend less time on a search. Thirdly, LTU leads to financial tension affecting the whole household and results in instability and uncertainty about the future.

The paper by Desiere et al. (2021) observes some models for predicting the LTU probability used by employment centers in different countries. All the models were applied on an individual-level data and covered information about the history of employment, personal characteristics and local labor markets.

Lots of models were based on statistical modelling like logit regression (Australia, USA, Austria, etc.) or probit regressions (Ireland). Different types of input variables were used that could be split into 4 types: socio-demographic factors (age, gender), motivational factors (expected salary, job-search behavior), job readiness factors (education, skills, experience), and opportunities (local markets information). Not all the described models used every group of variables but specific combinations of them. The accuracy score exceeded 0.6 but mostly was not over 0.8.

The significant advantage of such models is their results' interpretability, which most of the machine learning models cannot provide. This is an important feature as it gives a deeper understanding of a problem's causes.

On the other hand, machine learning models allow for the use of a larger set of variables (Haug, 2023) and give more accurate forecast (as they take into account non-linearity among variables). As the amount of data and its accessibility grow, there is the demand for using such models in a sphere of social policy. Those models were also used to predict the LTU probability. For example, random forest was applied in Belgium on data that included not only some standard set of variables but also a person's activity on an employment center's website. That enabled to estimate how intensive the job search process was.

Zhao (2020) estimates various models (logit regression, random forest and ensemble of decision trees) to predict individual hazard of a long-term unemployment. The author uses 171 determinants and evaluates contribution of a particular determinant by SHAP method (SHapley Additive exPlanations). XGBoost turns out to have the most accurate predictions with accuracy equals to 80%. SHAP method has shown that the most significant variables are “entry count to PES (Public Employment Services)” (lower risk of LTU), “experienced LTU before” (higher risk of LTU) and others. The first result is rather obvious, as people who are interested in finding a job quickly will visit an employment center more frequently. However, the last result is rather controversial. It is a common idea that the long-term unemployed lose some specific skills that results in an even higher risk of LTU. The author says that those findings demand further research.

Discussed papers solved classification problem so that the unemployment duration itself was not a target variable. Moreover, classification models do not consider data censorship that could have weakened forecast accuracy. Survival analysis enables overcoming problems mentioned above. Boškoski et al. (2021) investigates patterns of unemployment duration in Slovenia from 2011 to 2020 using survival analysis with machine learning models. The authors estimated models to predict duration variable. They used accelerated failure time model (AFT) as a base one but put neural network on the top of it to take into account all the potential non-linear dependencies among variables. They also used Bayesian methods to get an estimation of a prediction distribution instead of a point estimation. It gave a clearer picture about the potential hazards. Also, it allowed for estimating the probability of the LTU (longer than 180 days) and comparing prediction accuracy of the suggested model to previous ones.

The authors used train dataset that covers 12 months period and made predictions on the following 6 months. The dataset consists of the multiple categorical variables that were transformed to sets of dummy variables. The accuracy score of a classification model achieved 75.6%.

To sum up, models that predict hazards of LTU and its duration have only recently gained attention. Employment centers require those models to define the most vulnerable groups of people who seek some additional assistance to find a job. Most of the papers estimated classification models so that they get predictions that unemployment duration would exceed some certain period of time (that is defined by economic conditions and historic data). Even though those models are quite popular they have some considerable drawbacks. Firstly, they do not solve the problem of censored data points that worsens their predictive power. Secondly, classification models discriminate some groups of people (for example, they predict higher hazard for elderly people than is the case) that could affect efficiency of job search assistance (Zhao, 2020). Thirdly, classification models are rigid as they use a predefined value to measure LTU. Changing economic conditions affect the normal duration of unemployment and the definition of LTU. However, classification models are not that flexible as those that predict duration of unemployment.

Despite all these drawbacks, just a few papers favored duration models to predict unemployment duration but not classification ones. Even fewer of them used machine learning methods to strengthen predictive power. This paper constructs models to predict unemployment duration itself using duration analysis.

Moreover, some of them are based on machine learning techniques to capture all the non-linear dependencies among variables.

### 3. Methodology

Data censorship is an important feature of durational data. It is not always known whether a person has found a job at the time of closing his profile at an employment center because reasons for closure could vary. Those profiles contain information only about the minimal duration of a job search but not an exact one.

Survival analysis is a class of statistical methods to investigate time until some events that includes censored observations. This class introduces some new terms. The first one is the survival function which is the probability that a person searches for a job longer than time  $t$ .

$$S(t) = P(T > t) = 1 - F(t), \quad (1)$$

where  $T$  is a random variable that stands for job search duration and  $F(t)$  is the distribution function.

Kaplan–Meier curve is an empirical estimation of the survival function. It is a non-parametric method so that it does not show how the survival probability depends on various factors. The formula for estimation is as follows:

$$\widehat{S}(t) = \frac{\prod_{t_i: t_i < t} (n_{t_i} - d_{t_i})}{n_{t_i}}, \quad (2)$$

where  $n_{t_i}$  is the number of people registered at an employment center at time  $t_i$ ;  $d_{t_i}$  is the number of people who find a job at time  $t_i$ .

The second one is the hazard function which is the probability that a person finds a job at time  $t$  given that has not done so before:

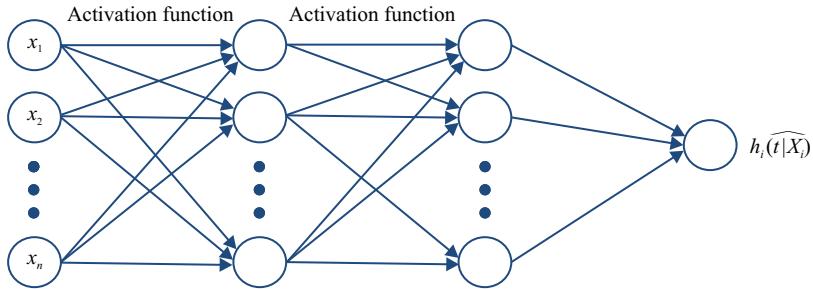
$$\lambda(t) = \frac{f(t)}{S(t)}. \quad (3)$$

One of the most popular survival regression models is Cox proportional hazards regression (Cox, 1972). The model states that there is a baseline hazard function and factors' change moves this function proportionally up or down. The model has the form:

$$\lambda(t | X_i) = \lambda_0(t) \times \exp(X_i \times \beta), \quad (4)$$

where  $X_i$  is the set of independent variables for subject  $i$ ;  $\lambda_0(t)$  is the baseline hazard function and  $\lambda(t | X_i)$  is the hazard function for subject  $i$  at time  $t$ .

The model is also used for estimating predictions, however, it does not predict the exact survival time but hazard function values at every time  $t$ . Hence it is possible to come to conclusion about the survival time, exploring predicted survival function. Moreover, it gives understanding about the probability that search time will exceed some certain amount of time (probability of LTU, in other words) that enables comparing this model to those estimated in other research.



**Fig. 1.** Schematic structure of Cox-Time neural network.

Source: Compiled by the author.

### 3.1. Machine learning models

The Cox model is not flexible, as its form is rigid. Machine learning models enable to capture more complicated dependencies among variables and make more accurate predictions. Kvamme et al. (2019) describes a modification of proportional hazards model based on neural network that is called Cox-Time model. The form of the model is as follows:

$$\lambda(t|X_i) = \lambda_0(t) \times \exp(g(t, X_i)), \quad (5)$$

where  $g(\cdot)$  is some function that is defined by neural network.

Neural network represents a structure of multiple layers of nodes that are connected among each other by edges (see Fig. 1). Those nodes are some real numbers, and edges are represented by some non-linear functions that are called activation function. The latter enables considering all the non-obvious connections among variables. In the paper ReLU (Rectified Linear Unit) is used as an activation function:

$$f(x) = \max(0, x). \quad (6)$$

Neural networks have input and output layers and some hidden ones that are aggregations of neurons. Given data observations are transferred as an input layer, then it is transformed by passing through hidden layers straight up to the output layer. In case of predictive models output layer is estimated predictions.

All the nodes have some weights that are estimated by adaptive minimizing loss function. Loss function shows how much estimated predictions are different from real data points. Adaptive training means that training dataset is divided into two parts: dataset for estimation models' parameters and validation dataset. A model predicts validation data points by iteratively updating its parameters. If loss function hasn't been changing during a few last epochs, the process stops.

Usually, loss function has the form of mean squared error, however, the following form is used for estimation Cox-Time model:

$$\text{Loss} = \frac{1}{n} \sum_{i:D_i=1} \log \left( \sum_{j:j \in R_i} \exp[g(T_j, X_j) - g(T_i, X_i)] \right), \quad (7)$$

where  $R_i$  is the restricted sample among all the subjects for which  $T_j \geq T_i$ .

The advantage of Cox-Time model is that  $t$  is used as one more explanatory variable that allows hazard function to be non-proportionate. It makes predictions much more flexible, however, it could lead to overfitting when a model shows good results on a training dataset while making low quality prediction on a test dataset. In this research dropout technique was used as a mean to prevent overfitting that is considered to be an easy and computationally effective way of regularization. Dropping out some random neurons enables to lower prediction error. The fraction of neurons that is going to be shut down refers to dropout rate.

The described model gives quite accurate predictions that consist of two parts: baseline hazard function and non-proportionate impact of explanatory covariates.

### 3.2. Prediction quality

There are multiple metrics to evaluate prediction accuracy of survival models. The most common ones are C-index and Brier score. C-index ranges from 0 to 1, with 1 indicating that a model's predictions are the most accurate. The index challenges the ability of a model to provide a reliable survival ranking. So that C-index would be high even if a model predicts higher (or lower) hazards than the actual ones for all the subjects proportionally.

The Brier index (Graf et al., 1999) compares predicted values with actual ones. The index is similar to mean squared error, however calculated for every time  $t$ . The index considers censored data and has the following form:

$$BS(t) = \frac{1}{n} \sum_{1 \leq i \leq n} w_i(t) (I(T_i > t) - \widehat{S(t|X_i)})^2, \quad (8)$$

where  $I(T_i > t)$  is the indicator function showing that the job search duration of a person  $i$  exceeds time  $t$ ;  $\widehat{S(t|X_i)}$  is estimated survival function for subject  $i$  at time  $t$ ;  $w_i(t)$  is the weight of subject  $i$  at time  $t$ . The more weight is given to non-censored events.

The integrated Brier score is used to estimate a model performance for the whole-time range and is calculated as follows:

$$IBS = \frac{1}{t_{max}} \int_0^{t_{max}} BS(s) dt, \quad (9)$$

where  $t_{max}$  is the maximum search time.

In the paper C-index is used as a classic tool to challenge model's predictions.

## 4. Data

In the paper, data of the Research Development Infrastructure (RDI)<sup>1</sup> was used. Data contains information about registered jobseekers in Russia from 2017 until 2022. All the missing observations were omitted. That led to exclusion of the whole regions from the research. Krasnoyarsk Krai and Belgorod Oblast are among those regions. Moreover, those unemployed younger than 18 and older than 61 were also excluded from the research, as their inclusion could cause biased results.

---

<sup>1</sup> RDI website: <https://data.rsci.science/about/> (in Russian).

As a result, 13 covariates were left where 2 variables are categorical ones, 2 are numerical, and others are dummy. All the variables can be divided into 4 groups (personal characteristics, marital status, local labour market and reservation wage) based on Lancaster and Nickell (1980). Personal characteristics' variables contain information about gender, sex, education and others. Marital status variables describe information about family of an unemployed. Dummy regional variables were introduced in order to capture differences in local labour markets. The last group of variables refers to reservation wage (a minimal wage rate at which the unemployed is ready to accept an offer). However, reservation wage itself cannot properly explain the difference in job search time, but it needs to be considered with salary that employers are ready to pay to a particular worker. Therefore, two dummy variables were generated. The first one indicates if a jobseeker has high salary expectations (expected salary is much higher than the one at the previous job) and the second one indicates if salary expectations are low (expected salary is lower than the one at the previous job).

Foley (1997) shows that reasons for dismissal are also crucial when modelling job search patterns. People who resigned of one's own accord spend more time on the search. So that dummy variable that shows if resignation was voluntary or not was added to the set of covariates. All the variables and their detailed descriptions are presented in Appendix A.

At the end, one-hot encoding was used as a step in data processing to transform all the categorical variables ("education level" and "region") into sets of dummy variables. "Secondary education" and "Nizhny Novgorod" were taken as reference categories. Also, all the numeric variables were centralized and normalized to achieve better prediction accuracy.

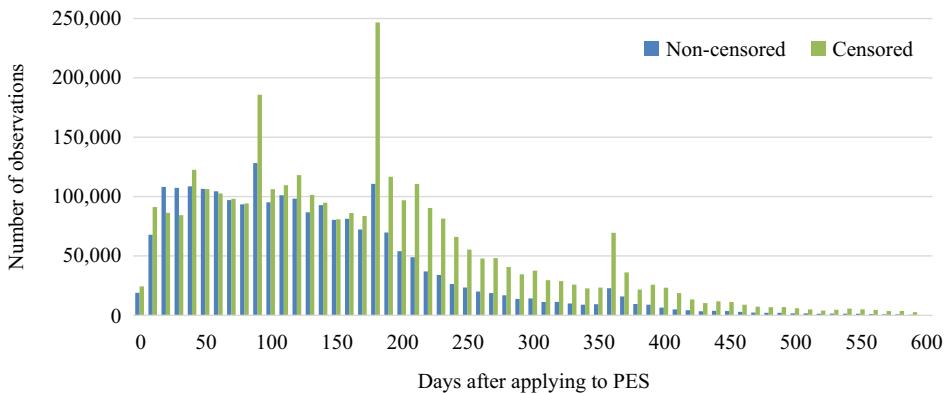
Unemployment duration was used as a target variable. Some observations do not exhibit information about the exact duration of a job search but only about the minimal one. Those observations are right-censored, and they should be treated differently from the non-censored observations. Data points are censored if an unemployed person stops visiting PES because of various reasons other than employment.<sup>2</sup> So that the job search time and information about censorship should be taken as dependent variables.

## 5. Descriptive analysis

Final dataset contains 5,608,597 data points with 3,312,188 being right-censored. In Fig. 2 there is target variable distribution for censored events and non-censored ones. Non-censored events refer to people whose personnel files were closed due to their employment. Censored observations occur if a reason was other than employment. "Pension assignment," "persistent absenteeism," "moving to another city" are among the most frequent ones.

Personnel files close because of employment earlier than the other reasons (in 139 days after applying to PES vs in 178 days). The number of closed files increases abruptly in 3 months, 6 months and 1 year after applying that is observed

<sup>2</sup> Federal Law of the Russian Federation No. 565-FZ of 12.12.2023 "About employment in the Russian Federation." [https://www.consultant.ru/document/cons\\_doc\\_LAW\\_60/9c4a1bccf20efb7bf09fab1d03518acc\\_c70ab30c/](https://www.consultant.ru/document/cons_doc_LAW_60/9c4a1bccf20efb7bf09fab1d03518acc_c70ab30c/) (in Russian).



**Fig. 2.** Unemployment duration distribution.

Source: Author's calculations.

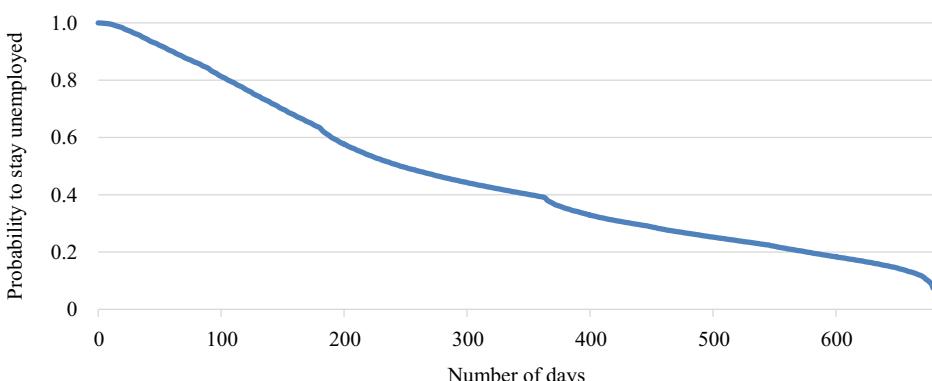
better for censored events. It could be due to the fact that unemployment benefits are paid over 3, 6 and 12 months based on individual circumstances. Probably most people are ceasing to visit employment center after that without notice.

Some descriptive statistics for covariates are presented in Appendix B. The average unemployed is a 40-year-old male with vocational education who resigned voluntary from his last job. Moreover, the tables show that some groups of the unemployed are scarcely represented (such as people without job experience, divorced, etc.). However, there are no reasons for these groups differ from majority in terms of job search duration.

Kaplan–Meier curves were estimated as a part of descriptive analysis. The same was done in the paper by Maigur (2023) on the same dataset, and here only some important results would be reported.

Kaplan–Maier curve is an empirical estimation of survival function. There is a curve estimated for the whole dataset in Fig. 3. The curve is restricted by 680 days as just a small share of the unemployed has been in a search for that long.

The curve has a stable negative slope and the probability to search for a job for more than a year, for example, consists of about 40%. Some patterns described above are noticed here as well: the curve drops at the points 3, 6 and 12 months.



**Fig. 3.** Kaplan–Meier curve.

Source: Maigur (2023).

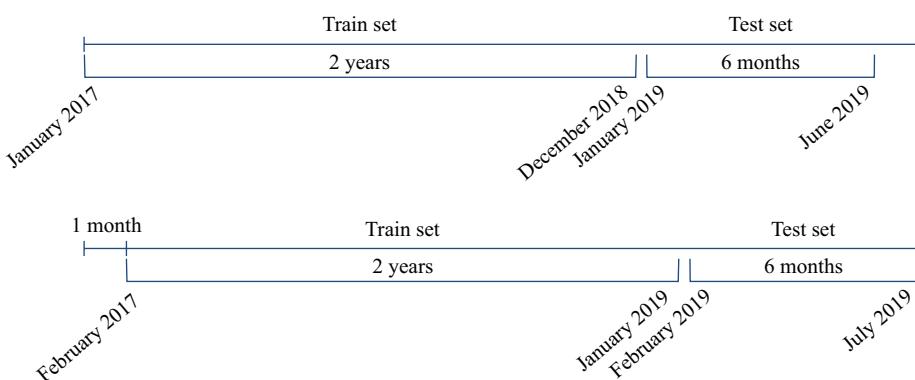
Such curves were also built for different groups of people to see if various characteristics have an impact on job search time. It turned out that it is more challenging to find a job for people over 50 years old and for people with just primary education. At the same time, it almost makes no difference in terms of unemployment duration for men and women and for other age groups.

## 6. Model specification

Proportional hazards and Cox-Time models were used to make unemployment duration predictions. Models were estimated on train datasets and predictions were made on test datasets. Boškoski et al. (2021) showed the method of consecutive predictions when 1-year period was used as a train set and following 6-month period as a test one. Then those sets were shifted for a certain period forward, and the procedure was repeated multiple times. In the current study this method was adopted to make train sets cover a 2-year period in order to capture long-term unemployment (as the longest job search duration lasted for almost 23 months). The longevity of test sets was initially set to 6-month, and sets were shifted 1 month forward each time.

In Fig. 4 an example of two pairs of train and test sets are presented. Firstly, a model is trained on a set that includes data from January 2017 until December 2018. Then a model performance is validated on a set covering the period from January 2019 till June 2019. After that, both train and test sets are shifted by one month forward (train set: from February 2017 till January 2018; test set: from February 2019 till July 2019), and model is trained and tested on new sets. The procedure was consequently repeated till the test set covered June 2021. This procedure provides static forecasts but enables to see how prediction quality changes over time.

All of this has enabled to consider how predictions accuracy has been changing over time. Moreover, 1-month period was also investigated as a test set. The longevity of a job search depends on various factors that also include some economic conditions and conjuncture. Therefore, long-term predictions could noticeably lose to short-term ones. Using different testing periods, it is possible to define the best strategy of implementation such model by PES.



**Fig. 4.** The scheme of time-series train-test split.

Source: Compiled by the author.

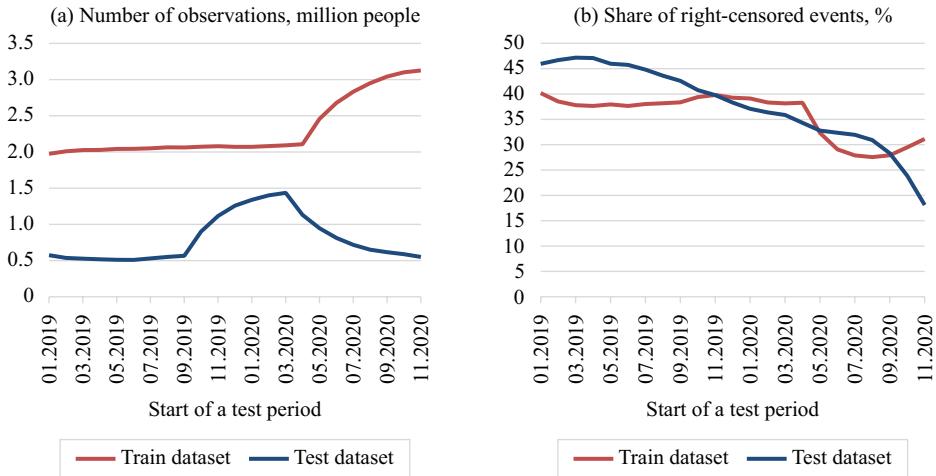


Fig. 5. Descriptive statistics for train and test sets.

Source: Author's calculations.

In Fig. 5a there is a number of observations in train and test sets (with test dataset being 6-month long), in Fig. 5b there is a share of right-censored events in those sets. The number of observations is growing during pandemic period. It could result in a drop in predictive power on test sets covering the first months of the pandemic as the model would not cover enough data points referring to some new economic conditions. At the same share of non-censored events is constantly decreasing, reaching its minimum at the latest available period. That is an obvious finding, as for the more recent observations there is less share of events.

The nonparametric proportional hazards model was used in the paper so that there was no need for choosing any parameters' sets. Coefficients of the model were estimated by maximizing the Cox likelihood function.

Parameter set for the Cox-Time model represents activation function, number of layers, number of nodes at each layer and dropout rate. All the tested combinations are presented in Appendix C.

The hyperparameter fit was made on the fifth part of the initial dataset. To reveal the best hyperparameter set, 5-folds cross-validated grid-search over a parameter was used. Therefore, this dataset was randomly split into 5 folds, a model with some combination of hyperparameters is trained on 4 folds and then performance is computed on the fifth part. The procedure is repeated 5 times with test set being one of the previously split folds. A final performance is averaged over the loop. The models' predictions with different parameters were compared by the values of C-index and time of estimation.

The problem that is common for neural network estimation is computational complexity that makes all the estimations extremely time-consuming. This point is crucial for the research because multiple neural network models for the various time periods need to be built instead of just one.

A model with 6 layers and 128 nodes at each one has reached the highest quality of predictions, however, a model with just 2 layers and 32 nodes did quite accurate predictions and had a great advantage in terms of the time of algorithm execution time. So that this architecture was chosen as a final one.

Also, various dropout rates were tested. It was found that a comparatively high rate of dropout leads to underfitting as the model does not have enough nodes for a good fit, while a low number of shut down nodes results in overfitting. Therefore, 10% of dropout seems to be the most effective way of regularization that was proven by validation.

The final step was to choose activation function out of ReLU, Sigmoid, Softmax and Tanh. All of them are quite widespread and actively used by researchers. Our experiments on real data showed that it makes almost no differences between using ReLU or Tanh functions, however, the first one is better in terms of execution time. Sigmoid and Softmax functions led to considerably lower prediction quality and were left out. So ReLU function turned out to be the optimal one following this prediction quality and execution time all together.

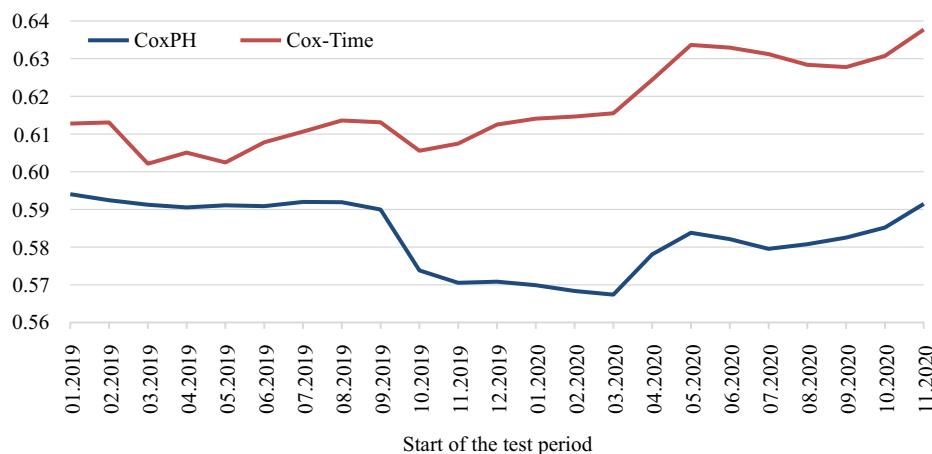
Hence, the final model architecture presents 2 layers of 32 nodes at each one with 10% dropout rate and ReLU activation function. The final model was trained and tested multiple times on the various preprocessed datasets, where preprocessing included one-hot encoding and numeric variables' standardization.

## 7. Results

### 7.1. C-index

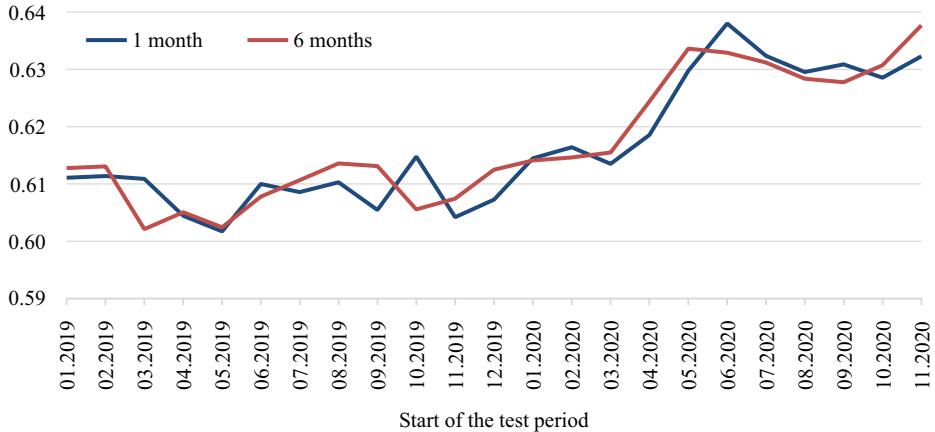
C-index was used to test prediction quality of two models. C-index shows how good a model predicts relative hazard of an event. C-index ranges from 0 to 1 while it is stated that C-index estimated on a real-data predictions mostly is not over 0.75.

In Fig. 6 index values are presented for both models. The index values have not exceeded 0.64. Index follows some same patterns for both the Cox-Time and Cox proportional hazards models: it is getting lower during the COVID-19 pandemic which had a huge impact on local labour markets. However, the index for predictions made by the proportional hazards model is lower than those made



**Fig. 6.** C-index values for predictions made by the Cox proportional hazards (CoxPH) and Cox-Time models.

Source: Author's calculations.



**Fig. 7.** C-index values for Cox-Time model predictions on 1-months and 6-months testing periods.

Source: Author's calculations.

by the Cox-Time model. On the other hand, it is quite consistent over time and ranges from 0.58 to 0.59.

The Cox-Time model considers non-linear dependencies and so should predict better. Moreover, model specification with non-proportionate hazard is used in the research. Kvamme et al. (2019) proves that using the Cox-Time model is not beneficial in terms of predicting relative hazards but absolute values. Nevertheless, C-index for the Cox-Time model is higher than for the proportional hazards model. C-index values for the Cox-Time model are more volatile over the entire period of time than for the proportional hazards model, but there is no such dramatic drop in the index during the pandemic.

So the Cox-Time model is relevant in terms of both better predictions while there is a stable economic situation and faster adaptation to the new economic conditions. The predictions' quality could become even higher if shorten test dataset, as it will provide more homogeneous economic conjuncture among train and test data.

Therefore, the Cox-Time model predictions were made on 1-month periods to challenge if the model predicts better short term. The values for C-index are presented in Fig. 7. It is clear that a shorter prediction period is not sustainably advantageous. As the model considers complicated connections among variables, it is able to keep up high prediction accuracy for long. This means that it is not necessary for a PES to constantly update a model with fresh data to maintain prediction quality at a decent level that can save some financial resources and increase functional efficiency.

Despite the fact that predictions made on different periods are of quite the same quality, it is still possible that different groups of variables contribute to this quality. In order to get more insights about patterns of job-search time, some information about factors' impact on unemployment duration should be acquired.

## 7.2. Variables contribution

It is crucial to understand the degree of variables contribution for successful implementation models' prediction by PES. Employment agencies should clearly

see which groups are the most vulnerable in order to come up with the effective employment programs. The great advantage of statistical models is their interpretability.

In the study the proportional hazards model was used as a statistical tool. The same one was also described in the paper by Maigur (2023) where detailed descriptive analysis was provided. The paper states that the most significant variables that have positive impact on a job search duration are pre-retirement age, primary education, voluntary dismissal and other, while master's degree, gender (male), relatively low salary expectations, and bachelor's degree have the most noticeable negative impact. Also, regional variables are of a major influence, with Novosibirsk Oblast being a region with the shortest search time and Khabarovsk Krai with the longest.

As it was noted above, the Cox-Time model results in higher prediction accuracy. However, machine learning models are mostly black box so that it is impossible to say what specific factors most contributed to the results. The following method was applied to find out what factors are the most significant for predictions: variables were one by one excluded from covariates set, and C-index values were compared for estimated models. The models were estimated not for a whole dataset but for an instance of training dataset starting from February 2019 and the following 6-months and 1-month testing datasets.

The results are presented in Table 1. The feature of the considerable effect for predictions is region: C-index drop by more than 0.06 in comparison to the initial model. It refers to both short- and long-term predictions.

All other variables also contribute to forecast but less significantly. For example, for 6-month testing period, the longevity of the last year's experience, and divorce and single parent records matter. While for a shorter period, education level and large family records are more important. Interestingly, the results of predictions on 6-months period were almost unchanged after the exclusion of those variables.

Including the variable "Pre-retirement age" almost does not affect forecast. That could be explained by the fact that variable "Age" itself is more informative for the model. Also, "No job experience" is of low impact that is mostly affected by the fact that this group of people is scarcely presented in the studied dataset.

**Table 1**  
C-index values for models estimated without one of the variables.

Excluded variable	1 month	6 months
Region	0.57	0.57
Large family	0.59	0.61
Education	0.59	0.61
Last year experience	0.60	0.59
Age	0.60	0.61
Resigned voluntary	0.60	0.60
High salary expectations	0.60	0.61
Single parent	0.60	0.60
Gender	0.60	0.61
Divorced	0.60	0.60
Low salary expectations	0.60	0.61
No job experience	0.61	0.61
Pre-retirement age	0.61	0.61

Source: Author's calculations.

To sum up, it is obvious that job-search time is heterogeneous among different regional labour markets that is confirmed by variables' contribution analysis made for both proportional hazard and Cox-Time models. Moreover, local labor markets contribute more to the accuracy of predictions, than any other variables. Personal characteristics also matter but to a lesser extent. For linear model age, education and salary expectations are factors of the most significance while family characteristics and length of last year experience are more crucial for neural network (for 6-months test period). Also, the Cox-Time model takes into account all the non-linearity among variables, therefore using the variable "pre-retirement" both with the variable "age" is excessive while it is one of the most impactful variables for the proportional hazards model.

## **8. Conclusion and discussion**

In this paper models for predicting the unemployment duration were estimated. The empirical analysis was based on data from employment centers of the Russian Federation from 2017 to the first half of 2021. The data is an aggregation of personnel files of the unemployed, which contain the information about the dates of opening and closing of personnel files, as well as some personal characteristics. Personnel files can be closed due to various reasons, including employment, relocation, long-term absence, etc. For observations that were closed for reasons other than employment, it is not possible to determine the exact duration of the job search, only its minimum duration. Observations of this type are defined as right-censored events, and in the analysis, they must be correctly taken into account, which is possible with methods of survival analysis. These methods cover a wide range of statistical models (parametric and nonparametric ones), as well as machine learning models, some of which were applied during the analysis.

In the paper two models were used: the Cox proportional hazards model and the disproportionate hazards model (the Cox-Time model). Both models take into account censored data, but the second model also removes the proportional hazards constraint and takes into account complex nonlinear relationships among variables. Models were estimated on a wide range of covariates that includes some personal characteristics of unemployed people, information about family status, about wage expectations as well as about local labor markets. The dataset was split into 23 pairs of train and test sets where the former cover a 2-year period and the latter take the following 6 months. A model was estimated on a train set and then predictions were made on a test set that enabled to mitigate an effect of economic shocks on prediction accuracy.

Prediction accuracy was estimated using C-index. According to the index values, models' forecasts show similar patterns: the forecast worsens during the coronavirus pandemic, which has had a significant impact on the labour market, the proportional hazards model gives a fairly stable result (except during times of COVID-19), and the index ranges from approximately 0.58 to 0.59.

The C-index for the Cox-Time model is higher. The index value is more volatile over the entire period, but such a dramatic drop in the index during the pandemic is not observed. The maximum index value (0.64) is achieved in the post-coronavirus period.

Additionally, for the Cox-Time model prediction accuracy was estimated to 1-month test set that showed that the model hits quite the same accuracy score no matter the length of test set. Hence, the model does not require constant update of the coefficients in order to sustain valid forecast.

Feature importance analysis for the Cox-Time model was also conducted that demonstrated that the most significant factor for unemployment duration is the region of a job search. Regional labor markets in Russia are so diverse that this information alone says a lot about job search patterns. The least important factors in a 6-month test period are those describing family status (divorced or single parent).

To sum up, the introduction and active use of Cox-Time model forecasts in employment centers will make it possible to identify a group of people who need additional assistance in finding a job. This, in turn, will lead to optimization of the costs of employment centers as well as a reduction in the average time to search for a job. However, revision of some aspects of the analysis in the future could potentially improve the quality of the models.

First, the data set used in this paper mainly covers categorical characteristics of the unemployed. Using methods to work directly with this type of data (Boškoski et al., 2021) would improve the quality of the forecast.

Secondly, the quality of the forecast was assessed using C-index referring to the relative value of the duration forecast instead of absolute one. That is, the index demonstrates high forecast quality for models that equally overestimate (or underestimate) the hazards for all individuals. Focusing on quality metrics that compare predicted values with actual values will allow identifying models with a high-quality forecast more accurately.

## 9. Policy implication

The Cox-Time model demonstrated high prediction quality on individual-level data. It seems that the elaborated model could be used in two ways.

Firstly, employment centers could easily implement the model into their daily activity. It is assumed that such a type of model does not demand high technological capacities, and its predictions will enable employment centers to define if a registered person is at high risk of long-term unemployment and needs some personal assistantship while searching for a job. This will optimize operational processes of these centers and contribute to more effective allocation of financial resources.

Secondly, this model could be used by policymakers. If there is persistent high unemployment rate, there is an urgent need for developing measures for stabilization. It was demonstrated that quality prediction of Cox-Time model is stable over time and does not fluctuate much under massive economic events. This model could become a tool for searching for the most vulnerable group of people among all the registered unemployed at a particular time, and based on this knowledge (their personal characteristics, family status and region of a job search) it is possible to formulate the actions that will lead to lowering of unemployment.

Hence there are multiple ways of using the Cox-Time model, each of which will contribute to reducing excessive unemployment.

## References

- Boškoski, P., Perne, M., Rameša, M., & Boshkoska, B. M. (2021). Variational Bayes survival analysis for unemployment modelling. *Knowledge-Based Systems*, 229, 107335. <https://doi.org/10.1016/j.knosys.2021.107335>
- Bover, O., Arellano, M., & Bentolila, S. (2002). Unemployment duration, benefit duration and the business cycle. *Economic Journal*, 112(479), 223–265. <https://doi.org/10.1111/1468-0297.00034>
- Bratberg, E., & Nilsen, O. A. (2000). Transitions from school to work and the early labour market experience. *Oxford Bulletin of Economics and Statistics*, 62(s1), 909–929. <https://doi.org/10.1111/1468-0084.0620s1909>
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- Denisova, I. A., Donetskij, A. M., Kolesnikova, O. A., Fedchenko, A. A., & Lyadova, N. I. (2003). Long stay in the register of unemployed: Low level of education, an unfortunate combination of circumstances, or something else?. In: *Social policy: The realities of the 21<sup>st</sup> century* (iss. 1, pp. 73–102). Moscow: Signal (in Russian).
- Desiere, S., Langenbucher, K., & Struyven, L. (2019). Statistical profiling in public employment services: An international comparison. *OECD Social, Employment and Migration Working Papers*, No. 224. <https://doi.org/10.1787/b5e5f16e-en>
- Foley, M. C. (1997). Determinants of unemployment duration in Russia. *Center Discussion Paper*, No. 779. Economic Growth Center, Yale University.
- Giltman, M., Merzlyakova, A., & Antosik, L. (2022). Exit from registered unemployment: Estimating the impact of individual characteristics. *Public Administration Issues*, 1, 193–219 (in Russian). <https://doi.org/10.17323/1999-5431-2022-0-1-193-219>
- Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, 18(17–18), 2529–2545. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990915/30\)18:17/18<2529::AID-SIM274>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5)
- Grogan, L., & Van Den Berg, G. J. (2001). The duration of unemployment in Russia. *Journal of Population Economics*, 14(3), 549–568. <https://doi.org/10.1007/s001480000029>
- Haug, K. B. (2023). Structuring the scattered literature on algorithmic profiling in the case of unemployment through a systematic literature review. *International Journal of Sociology and Social Policy*, 43(5/6), 454–472. <https://doi.org/10.1108/IJSSP-03-2022-0085>
- Jolianis, Adrimas, Bachtiar, N., & Muharja, F. (2020). Unemployment duration of educated workers in the provinces of Indonesia: A cross sectional analysis from labor supply perspectives. *Journal of Applied Economic Sciences*, 15(1), 97–105.
- Kooreman, P., & Ridder, G. (1983). The effects of age and unemployment percentage on the duration of unemployment: Evidence from aggregate data. *European Economic Review*, 20 (1–3), 41–57. [https://doi.org/10.1016/0014-2921\(83\)90056-9](https://doi.org/10.1016/0014-2921(83)90056-9)
- Kvamme, H., Borgan, Ø., & Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. *Journal of Machine Learning Research*, 20(129), 1–30.
- Lancaster, T., & Nickell, S. (1980). The analysis of re-employment probabilities for the unemployed. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 143(2), 141–152. <https://doi.org/10.2307/2981986>
- Luboyova, M., & van Ours, J. C. (1997). Unemployment dynamics and the restructuring of the Slovak unemployment benefits system. *European Economic Review*, 41 (3–5), 925–934. [https://doi.org/10.1016/S0014-2921\(97\)00050-0](https://doi.org/10.1016/S0014-2921(97)00050-0)
- Maigur, A. (2023). Regional analysis of registered unemployment duration. *Applied Econometrics*, 72, 100–120 (in Russian). <https://doi.org/10.22394/1993-7601-2023-72-100-120>
- Nivorozhkin, A. (2006). New estimates of the risk and duration of registered unemployment in urban Russia. *International Journal of Manpower*, 27(3), 274–289. <https://doi.org/10.1108/01437720610672176>
- Stern, S. (1989). Estimating a simultaneous search model. *Journal of Labor Economics*, 7(3), 348–369. <https://doi.org/10.1086/298212>
- Zhao, L. F. (2020). Data-driven approach for predicting and explaining the risk of long-term unemployment. *E3S Web of Conferences*, 214.01023. <https://doi.org/10.1051/e3sconf/202021401023>

## Appendix A

**Table A1**

Variable descriptions.

Group	Variable	Description
Personal characteristics	Gender	Variable is set to 1 if gender is female and to 0 otherwise.
	Age	Age of an unemployment at the time of registrations.
	Last year experience	Work experience for the last 12 months. Work experience is calculated in weeks. All values larger or equal to 53 are reported as 53.
	No job experience	Variable is set to 1 if an unemployed had no job experience is female and to 0 otherwise.
	Pre-retirement age	Variable is set to 1 if an unemployed if close to retirement age and to 0 otherwise.
	Resigned voluntary	Variable is set to 1 if the dismissal from the previous job was voluntary and to 0 otherwise.
Education		Variable can take values “Primary education”, “Secondary education”, “Lower post-secondary vocational education”, “Basic general education”, “Bachelor’s degree”, “Master’s degree”, “Specialist degree”, “Training of specialists of higher qualification”, “Other”.
		The highest level of education of a citizen is specified.
Family characteristics	Large family	Variable is set to 1 if an unemployed is a parent in a large family and to 0 otherwise.
	Divorced	Variable is set to 1 if an unemployed is divorced and to 0 otherwise.
	Single parent	Variable is set to 1 if an unemployed is a single parent and to 0 otherwise.
Local labour market demand	Region	Region of registration in an employment center (79 regions in total).
Salary factors	Large salary expectations	Variable is set to 1 if expected salary is much lower than one at the previous job and to 0 otherwise.
	Low salary expectations	Variable is set to 1 if expected salary is lower than the one at the previous job and to 0 otherwise.
	Pandemic period	Variable is set to 1 if job search was during pandemic period and to 0 otherwise.
Target variables	Unemployment duration	Unemployment duration (in days). It is calculated as the difference between date of registration at an unemployment center and date of deregistration.
	Employed	Variable is set to 1 if personnel file was closed due to employment and to 0 otherwise.

Source: Compiled by the author.

## Appendix B

**Table B1**

Descriptive statistics for numerical variables.

Variable	Mean	Std. dev.	Minimum	Median	Maximum
Unemployment duration	162	113	1	142	679
Age	40	10	18	39	61
Last year experience	26	21	0	29	53

Source: Author's calculations.

**Table B2**

Descriptive statistics for categorical variables.

Variable	Number of unique values	Mode	Frequency of the most frequent value, %
Employed	2	0	59
Gender	2	1	53
Education	9	Lower post-secondary vocational education	35
No job experience	2	0	99
Pre-retirement age	2	0	89
Large family	2	0	96
Divorced	2	0	98
Single parent	2	0	98
Large salary expectations	2	0	87
Low salary expectations	2	0	76
Resigned voluntary	2	1	66
Region	79	Moscow	7

*Source:* Author's calculations.

## Appendix C

**Table C1**

Subset of the hyperparameter space.

Hyperparameters	Values
Layers	{ <b>2</b> , 4, 6, 8}
Nodes	{ <b>32</b> , 64, 128, 256}
Dropout rate	{0, <b>0.1</b> , 0.5}
Activation function	{ReLU, Softmax, Sigmoid, Tanh}

*Note.* The selected parameters are highlighted in bold.*Source:* Author's calculations.