Kohonen maps to organize staff recruitment and study of workers’ absenteeism

V V Kukartsev¹,², O D Beletskaya¹, M O Fabrichkina¹, A Yu Telkov¹, N Y Belyaev¹, V S Tynchenko¹,² and A S Mikhalev¹

¹Siberian Federal University, 79, Svobodny pr., Krasnoyarsk, 660041, Russian Federation
²Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

E-mail: vlad_saa_2000@mail.ru

Abstract. As the economy transited to market relations, the problem of organizing personnel policy and staff recruitment becomes more acute. The article discusses the problem of worker absenteeism, as well as the ways to study it, identified on the basis of literature analysis. The paper proposes an approach to use Kohonen self-organizing maps to solve the problem. The authors conducted a study of the dependence of the workers’ absence duration in the workplace on their various attributes, such as age, distance from residence to work, transportation expenses and others. An open data set of a courier company was used to identify dependencies. As a result, three worker portraits were made and the main reasons for absence of each worker type were identified.

1. Introduction
In the modern world, most people aim to find successful job to satisfy their needs, as a result, the total number of laborers is growing, and now two important indicators should be considered: employment and unemployment level, which are one of the main characteristics of the country's economy. Labor is the main factor in the process of production of goods and services, which we consume and thereby support the country's economy. Therefore, workers who regularly perform their labor duties and consumers who regularly use the provided services ensure stability in the economy. However, such regularity is not always present. For instance, many employed citizens do not always fulfill their duties on time, and this is due to various reasons. In these cases, such phenomenon as worker absenteeism should be considered.

In general, worker absenteeism is a habitual pattern of absence from some duties. However, in this paper a more specific case is considered, which is applied in the field of labor-management relations and describes worker absenteeism as a value of the frequency and type of a person’s absence in the workplace. There are various excuses for the absence of a worker in their workplace, these excuses can be both reasonable and unreasonable. One of the most common good excuses is morbidity. According to the Rosptrebnadzor office of Krasnoyarsk Krai, only in March 2019, approximately 43 thousand cases of infectious diseases were registered [1]. Nevertheless, an unwellness situation can also be an example of absence without a reasonable excuse, for instance, when a worker did not appear in their workplace, but they also did not see a doctor.
The phenomenon of worker absenteeism is a complex problem that affects employees, employers, the workplace, as well as various social and economic factors [2]. Experts at the Falcongaze Analytical Center conducted a study on the reasons for the change of work by employees of Russian companies. According to the results of the study, the percentage of employees fired due to disruption of the work process in 2016 was 27% [3], and 16% in 2017 [4].

2. Literature review
The problem of worker absenteeism is not new and has already been addressed by many authors who have conducted various studies on the absence of employees in the workplace. These studies range from papers defining and classifying absenteeism [5, 6, 7] to papers suggesting a system for predicting absenteeism [8].

Most studies of absenteeism focus on the individual characteristics of workers. The papers of Gary Johns in 2010 [6] and 2011 [7] show that such factors like a worker’s personal qualities and their attitude to the work influence the frequency of workers appearing in the workplace. Factors affecting work attendance were also described by Geneviève Jourdain and Michel Vézina in their paper “How psychological stress in the workplace influences presenteeism propensity: A test of the demand–control–support model” [9]. The authors found that leadership support is an important factor. In the studies of various authors, one can find factors which seem to be irrelevant at first glance. For example, in one of the papers [10] the organization's attitude to the public or private sector was defined as one of the factors. In later studies, the dependence of worker absenteeism on social status, nationality and society was noted [11]. One of interesting factors is the distance from home to work, identified in a 2019 paper [12].

The studies often include large data sets, which complicates their manual processing. This necessitates the automation of the data analysis process, i.e., the use of mathematical methods. For instance, Gary Johns in his studies used such mathematical methods as the least squares and Student's t-test [13]. Also, Gary Johns and other authors often use meta-analysis in their papers. One of the studies includes using a method based on the Gioia methodology to analyze data [14]. In the same study, a qualitative approach presented in the form of interviewing is used in data analysis.

The scope of worker absenteeism is vast and has been studied and described by many authors. However, aforementioned studies in this area do not describe all the possible dependencies and factors affecting the absence of workers in their workplace. For this reason, it was decided to conduct this study and make portraits of workers depending on the duration of their absence from the workplace.

3. Methodology description
Kohonen map was chosen for this study. Kohonen map or Kohonen self-organizing map is a neural network using unsupervised learning which performs clustering and visualizes the multidimensional properties of objects on a two-dimensional map [15-17]. These two tasks were the main reasons for choosing this method, since in order to divide the initial data into subgroups, the method has to support clustering. The initial data are multidimensional, therefore, for their better perception, it is necessary to visualize the data, which Kohonen maps are used for. Moreover, the model is resistant to noisy data [18].

Kohonen map reflects the proximity of multidimensional feature vectors, therefore, objects whose feature vectors are close together are located in neighboring cells or go into one cell [19]. Kohonen maps are colored to analyze the parameters of similarity of objects.

In this study, the Deductor analytical platform, which is the basis for creating complete application solutions, is used to analyze data and build Kohonen maps.

4. Data set description
The data set “Absenteeism at work Data Set” [20], which is publicly available in the machine learning data repository, was chosen as the studied data. The data set was provided by the University of Nove de Julho (Universidade Nove de Julho – Postgraduate Program in Informatics and Knowledge
Management). It is based on absence data of courier company employees in Brazil from July 2007 to July 2010.

The data set consists of 740 records and includes 21 attributes: individual identification, reason for absence, month of absence, day of the week, seasons, transportation expenses, distance from residence to work, service time, age, workload average/day, hit target, disciplinary failures, education, children, drinking habit, smoking habit, pets, weight, height, body mass index, absence time in hours.

5. Adjusting parameters
Before working with the data set, it is necessary to process it to achieve better conditions for sampling. This process is called data normalization.

After normalizing the data, the training data set is configured. To do this, it is necessary to break the initial set into training and test subsets. The training subset includes entries containing the input and target output values. The test subset is used to validate the training step. The initial data set was randomly divided into training and test subsets with the ratio of 95% to 5%, respectively.

The next step involved configuring the parameters of the map, as well as the stop parameters of the learning algorithm. The error value was set to 0.05. The error value is the criterion for stopping the algorithm, it means if the error value calculated at one of the iterations of the algorithm is less than the set error value, then the algorithm stops running, and the training is considered completed. To avoid infinite loop, the limited number of epochs was set. Upon reaching this number, learning stops even if the error value was not reached. The value of 500 was chosen.

Also, two parameters must be set: learning rate at the beginning and end of training. The learning rate at the beginning of training is higher, since it is used in the rough adjustment phase, which is characterized by a rather large correction of the weights of the network neurons. During the fine-tuning phase, the correction value significantly decreases, therefore, the learning rate should be lower. The learning rate at the beginning of training was set to 0.3 and at the end of training to 0.005.

The learning algorithm was configured so the program itself would determine the number of clusters. The value of the “Significance Level” parameter was set to small for a smaller number of clusters.

After completing the steps to set up data processing, the next steps were training and visualizing the results.

6. Results
After the training was completed, the results visualized in the form of Kohonen maps were drawn, which are presented in figures 1 and 2. By the color shade, it is possible to determine the value of each of the features: lighter colors correspond to higher values. To determine it precisely, an example one of the features in more detail can be considered, for instance, the reason for the absence of a worker in the workplace. Figures 1 and 2 show that in the place of the second cluster there are cells of predominantly dark color, which means the second cluster includes workers who are absent mainly due to illness.

![Kohonen maps](image_url)

**Figure 1.** Kohonen maps for such features as transportation absence time in the workplace (a), expenses (b), reason for absence (c).
Figure 2. Kohonen maps for such features as day of the children (d), week (e), height (f), drinker habit (g), workload average/day (h), smoking habit (i), distance from residence to work (j), pets (k), hit target (l), age (m), weight (n), service time (o), month of absence (p), body mass index (q), education (r), season (s) and also formed clusters (t).
As a result, 3 clusters are formed:

- 1 – minimum absence duration.
- 0 – medium absence duration.
- 2 – maximum absence duration.

The cluster of minimum absence duration is the biggest with 57.6% of all the data, the next is the cluster of medium absence duration with 40.4% and the cluster of maximum absence duration with 2%.

Also, for each of the factors, the levels of their significance were analyzed, according to which the factors that had the greatest classification impact were found. The greater the impact made by a particular input attribute when classifying the output, the higher the level of significance. According to the results of the study, those factors were: absence time in the workplace, transportation costs and the presence of a child in the family.

As a result of the analysis of the drawn maps, the portraits of workers were made, based on the duration of their absence in the workplace.

### Table 1. Attributes’ description for each worker type.

| Attribute                  | Absence duration |
|----------------------------|------------------|
| Age                        | Minimum | Medium | Maximum |
| Weight (kg)                | 36       | 37     | 40      |
| Height (cm)                | 78       | 80     | 80      |
| Education                  | 172      | 172    | 177     |
| Drinking habit             | Yes      | Yes    | Yes     |
| Smoking habit              | No       | Yes, 50/50 probability | No |
| Children                   | 0 or 1 child | 1 child (most likely) or 2 children | 2-3 children |
| Service time (years)       | 12       | 13     | 13      |
| Distance from residence to work | Long     | Medium | Short   |
| Transportation expenses    | Medium   | High   | Medium  |
| Pets                       | 0 or 1 pet | 2-4 pets | 1 pet  |
| Main reason for absence    | Medical reason (blood donation, visiting dentist, medical examination, etc.) | Short illness, medical reason | Long illness |

7. Conclusion

Worker portraits based on the analysis of Kohonen maps can help enterprise managers predict the behavior of their employees and take the necessary measures to reduce the probability of absenteeism in the workplace. Since each worker type has different characteristics and has a different main absence reason in the workplace than other types, the strategies for reducing absenteeism should differ. In this case, the management can take the following measures:
• Provide transportation service for employees who live in remote areas or have high transportation expenses, which is highly important for workers of groups with a minimum and medium absence duration.
• Hold meetings to promote healthy lifestyle that will help increase employees' awareness of how to maintain and strengthen their health (especially relevant for the group of workers with maximum absence duration).
• Improve the workplace climate (relevant for all groups of workers).
• Improve and maintain stable psychological work environment (relevant for all groups of workers).

The problem of worker absenteeism exists and is present in all organizations, and the aforementioned measures can reduce the frequency of employees’ absence in the workplace.

References
[1] Socioeconomic status of Krasnoyarsk Krai from January to March of 2019 Russian Federal State Statistics Service Retrieved from: http://web.krasstat.gks.ru/doklad/12/doc.htm#05-1.3
[2] Workplace attendance and absenteeism The Royal Australasian College of Physicians Retrieved from: https://www.racp.edu.au/docs/default-source/faculties-library/workplace-attendance-and-absenteeism.pdf?sfvrsn=4
[3] Employees' reasons for job change in 2016 Falcongaze, Ltd Retrieved from: https://falcongaze.ru/pressroom/publications/research/dismissal-reasons-2016.html
[4] Employees' reasons for job change in 2017 Falcongaze, Ltd Retrieved from: https://falcongaze.ru/pressroom/publications/research/reasons-for-change-of-job-2017.html
[5] Johns G 2008 The SAGE Handbook of Organizational Behavior (London: SAGE Publications Ltd)
[6] Johns G 2010 Presenteeism in the workplace: A review and research agenda Journal of Organizational Behavior 31(4) 519-42
[7] Johns G 2011 Attendance dynamics at work: the antecedents and correlates of presenteeism, absenteeism, and productivity loss Journal of Occupational Health Psychology 16(4) 483
[8] Ferreira R, Martiniano A, Napolitano D, Romero M, De Oliveira Gatto D, Farias E and Sassi R 2018 Artificial neural network for websites classification with phishing characteristics Social Networking 7 97-109
[9] Jourdain G and Vezina M 2014 How psychological stress in the workplace influences presenteeism propensity: A test of the Demand–Control–Support model European Journal of Work and Organizational Psychology 23(4) 483-96
[10] Bergstrom G, Hagberg J, Busch H, Jensen I and Bjorklund C 2014 Prediction of sickness absenteeism, disability pension and sickness presenteeism among employees with back pain Journal of occupational rehabilitation 24(2) 278-86
[11] Russo S D, Miraglia M, Borgogni L and Johns G 2013 How time and perceptions of social context shape employee absenteeism trajectories Journal of Vocational Behavior 83(2) 209-17
[12] Ma L and Ye R 2019 Does daily commuting behavior matter to employee productivity? Journal of Transport Geography 76 130-41
[13] Edwards V L 2017 Examining absenteeism in the public and non-profit sectors International Journal of Organization Theory & Behavior 17(3) 293-310
[14] Ruhle S A and Suss S 2019 Presenteeism and Absenteeism at Work—an Analysis of Archetypes of Sickness Attendance Cultures Journal of Business and Psychology 1-15
[15] Kohonen T 2001 Self-organizing maps (New York: Springer)
[16] Tynchenko V S, Tynchenko V V, Bukhtoyarov V V, Kukartsev V V, Kukartsev V A and Eremeev D V 2019 Application of Kohonen self-organizing maps to the analysis of
enterprises’ employees certification results IOP Conference Series: Materials Science and Engineering 537(4) 042010

[17] Milov A V, Tynchenko V S, Kukartsev V V, Tynchenko V V and Bukhtoyarov V V 2018 Use of artificial neural networks to correct non-standard errors of measuring instruments when creating integral joints Journal of Physics: Conference Series 1118(1) 012037

[18] Tynchenko V S, Tynchenko V V, Bukhtoyarov V V, Tynchenko S V and Petrovskyi E A 2016 The multi-objective optimization of complex objects neural network models Indian Journal of Science and Technology 9(29) 99467

[19] Tynchenko V S, Petrovsky E A and Tynchenko V V 2016 The parallel genetic algorithm for construction of technological objects neural network models 2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM) 1-4

[20] Absenteeism at work Data Set UCI Machine Learning Repository Retrieved from: http://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work