

ICEST 2021

II International Conference on Economic and Social Trends for Sustainability of Modern Society

PERSONNEL AUDIT BY MACHINE LEARNING METHODS

M. L. Krichevsky (a), S. V. Dmitrieva (b)*

*Corresponding author

(a) State University of Aerospace Instrumentation, ul. Bolshaya Morskaya, 67, Saint-Petersburg, Russia,
mkrichevsky@mail.ru

(b) State University of Aerospace Instrumentation, ul. Bolshaya Morskaya, 67, Saint-Petersburg, Russia,
dsv949@yandex.ru

Abstract

In present-day economic conditions, a new management paradigm has emerged. According to this paradigm the main source of the organization's earned value is not material capital, but people who have knowledge and conditions for the implementation of the goals of the enterprise. Personnel audit in an organization is the most difficult element of personnel management. Essentially, audit is characterized by some activities, in terms of which assessments of the employees and the results of their activities are formed. When constructing such an assessment, it is necessary to identify to what extent the results of the contractor's work correspond to the desired goals and regulatory requirements. When conducting an audit of human capital, highly trained specialists are needed in the field of economics, human resources and management. The auditor must identify the key skills needed by employees to achieve the organization's core objectives, and has an interest in identifying employees whose knowledge and skills are used not in full. To form objective assessments of personnel during an audit, it is offered to use machine learning methods, the results of which are a kind of assistants to auditors when deciding on the quality of work of a particular performer. The methods of classification in a situation where employees need to be divided into specific categories according to the degree of their labor efficiency are considered in this article.

2357-1330 © 2021 Published by European Publisher.

Keywords: Personnel assessment, quality of the contractor's work



1. Introduction

HR audit (personnel audit) is a term used to research human resources in an organization. Personnel audit can have different purposes and benefits in accordance with the current needs of the organization. The subject of the audit is an objective independent review and assessment of the state of the staff, as well as proposed solutions to improve the efficiency of individuals, departments and the entire organization (Alpaydin, 2010). Personnel audit is a tool to improve the efficiency and effectiveness of an organization. HR audit is usually used in situations where it is necessary to assess the condition of employees and improve the functioning of the organization. Such an audit includes an assessment of the skills of employees within the firm and its organizational structure. It is an essential part of the experience of any company for its employees. After conducting a detailed personnel audit, the top management of the organization can answer questions about more balanced distribution of work among employees, identifying the most competent employees for specific positions and the best way to manage employee productivity (Azad & Moshkov, 2017).

The essence of personnel audit, like any other type of audit, boils down to a simple formulation: control, evaluation of the process and results. The form of control, its content is determined by the goals and objectives of the audit, as well as specific features of the state of the internal environment of an economic entity. Assessment of the organization's personnel is the most difficult link in personnel management. It represents special formalized measures, according to the results of which the employees themselves, their work and the result of their activity are assessed. The development of personnel audit methods and the formation of an assessment of the performance of employees are the goals of this article.

2. Problem Statement

For a high-quality HR audit, a well-designed audit program plays an important role. To do this, it is necessary to determine the list of documents to be checked. Currently, there is no legally established list of documents subject to verification as part of a personnel audit (Coppin, 2004). As a result of this, the auditor independently determines such a list, based on the specifics of the enterprise. Although various sources show lists of indicators for assessing the work of employees of different levels, for example, the head of an organization, line managers, heads of departments, rank-and-file performers, nevertheless, the authors do not know the aggregate estimates of the personnel audit. Given this circumstance, the purpose of the article mentioned above is transformed into the use of artificial intelligence methods to build an assessment of the performance of employees. As you know, an important part of artificial intelligence is machine learning (Ezghari, et al., 2017). The term "machine learning" refers to a technology that is defined in (Gao, et al., 2019, p. 105748): "Optimization of the performance criterion of a model using data and past experience". Taking into account that in recent years the traditional methods of data processing have been replaced by machine learning models, the audit problem posed in this work is reduced to the development of personnel assessment methods using such models.

3. Research Questions

The list of questions to be investigated in this work consists of the following set:

1. What characteristics of an employee (to the greatest extent) affect the efficiency of his work?
2. How to create a database of the organization's personnel to work with machine learning models?
3. How to evaluate the quality of the generated data?
4. What models of machine learning to choose for research in the task of personnel audit?
5. How to choose the best model?
6. How to evaluate the obtained result in terms of its suitability for the task at hand?
7. How to use the chosen model to solve the problem of personnel audit?

The answers to the questions posed constitute the content of this article.

4. Purpose of the Study

The purpose of the study is to analyze the methods of machine learning for solving the problem of personnel audit, the formation of the necessary database for use in the selected methods of machine learning, assess the quality of the solution obtained and demonstrate the application of the proposed method for personnel audit.

5. Research Methods

The main task of machine learning is to form an inference from a sample of data. It should be noted that, firstly, training requires effective algorithms to solve the optimization problem, as well as to store and process the huge amount of data that is usually available. Secondly, once the model has been learned, its presentation and algorithmic decision for inference must also be efficient. In some applications, the effectiveness of a learning or inference algorithm, namely its spatial and temporal complexity, can be just as important as its prediction accuracy. The name "machine learning" implies that the method analyzes the data, it itself, and not a person, finds the model. It is "training" as it looks like training. Thus, the information used in machine learning are training data.

5.1. Database creation

To create a database of examples used in machine learning for building a model, two approaches can be applied:

- collect real data;
- form a toy dataset.

Let's assume that we have no real data for the current study, so let's consider the second method. In machine learning, it is important to learn how to use such sets correctly, since learning algorithms based on real data is difficult and may fail (Kim & Hong, 2017). Toy datasets play a critical role in understanding how an algorithm works. Simple synthetic data sampling simplifies the situation in order to assess whether the algorithm has learned the necessary rule or not. It is difficult to make such an assessment using real data. In this case Monte Carlo method is used to sample synthetic data.

5.2. Preprocessing - Principal Component Method

Principal component (PM) method is one of the most important and powerful methods for data preprocessing (Personnel Audit, 2021). The purpose of introducing PC is to explain the variance of the observed data in terms of a linear combination of the original data. This method searches for new variables that are linear functions of the variables in the original dataset. The search for such new variables - the main components - is reduced to solving the problem of eigenvalues (eigenvectors) of the correlation matrix of the initial data.

PC method makes it possible to represent multi-feature objects, which in this work are understood as employees of the enterprise, in a space of lower dimensions, to identify between them the similarity or difference in a set of features, to form preliminary conclusions about the presence of groups (classes) of objects. PC method is somewhat similar to factor analysis, but there are certain differences between them.

5.3. Classification tools

For the task of personnel audit in this work, the methods of machine learning will be used, which give a solution in the form of discrete classes, for example, an employee, according to the degree of his labor efficiency, belongs to the first, second or third categories. Some of these methods are as follows:

- Tree Decision (Personnel-audit, 2021; Principal Components Analysis, 2021).
- Linear Discriminant Analysis (Ramsundar & Zadeh, 2018; Vapnik, 1998).
- Support Vector Machine (Yashkova & Timofeeva, 2019)
- The k-Nearest Neighbors (Zhang, et al., 2017).

6. Findings

When auditing personnel, the important characteristics that affect the quality of employees' work are the following: X1 is gender; X2 is age; X3 is education; X4 is work experience; X5 is initiative; X6 is motivation; X7 is effectiveness in achieving goals; X8 is knowledge of a foreign language. The output parameter is labour efficiency. Of course, the choice of such features has elements of subjectivity, but, according to the authors, it describes the situation in a personnel audit quite fully.

When forming the database, we will choose three efficiency classes and use Monte Carlo method to play out the characteristics of employees. In this case, the parameters are modelled from the following ranges: for X1 - (0; 1); for X2 - (22-50), years; X4 - (1-6), years; other signs - (1 -10), point. In addition, during the simulation, random noise was added to level out individual outliers.

Each of the classes contained 10 objects, so the complete simulated database of examples is a table consisting of 30 rows and 9 columns (8 columns are signs; the 9th column is the name of the class).

Table 1 shows a snippet of the raffled data.

Table 1. Simulated database of examples

| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | Class |
|----|------|------|-----|------|------|------|------|-------|
| 0 | 28.7 | 8.3 | 4.5 | 8.6 | 7.6 | 8.4 | 7.1 | First |
| 0 | 27.0 | 7.6 | 2.5 | 6.1 | 6.9 | 7.6 | 8.7 | First |
| 0 | 27.0 | 13.1 | 7.3 | 10.0 | 12.7 | 11.7 | 10.8 | First |
| 0 | 27.1 | 9.1 | 7.2 | 9.4 | 10.5 | 11.2 | 11.1 | First |
| 1 | 28.7 | 13.6 | 9.2 | 10.7 | 12.7 | 12.0 | 13.5 | First |
| 0 | 51.2 | 11.3 | 9.7 | 9.9 | 13.0 | 13.1 | 11.2 | Third |
| 1 | 44.9 | 0.5 | 0.2 | 2.0 | 1.0 | 4.0 | 3.3 | Third |
| 0 | 48.0 | 6.3 | 5.0 | 5.9 | 5.1 | 9.1 | 5.0 | Third |
| 0 | 51.7 | 3.3 | 4.6 | 4.4 | 5.9 | 3.2 | 5.4 | Third |
| 1 | 48.7 | 4.0 | 4.4 | 7.4 | 4.4 | 4.3 | 5.0 | Third |

In Figure 1, all the data played out are shown on the plane of the first two main components (MC), from which it can be seen that there is some overlap of classes. From Figure 1, it follows that the loss of information content in the transition to two MCs is about 14%, the value of which is quite acceptable with an insignificant amount of observations. In this case, it becomes possible to visualize data in a reduced two-dimensional space (instead of working in the original 8-dimensional space, where such visualization is impossible).

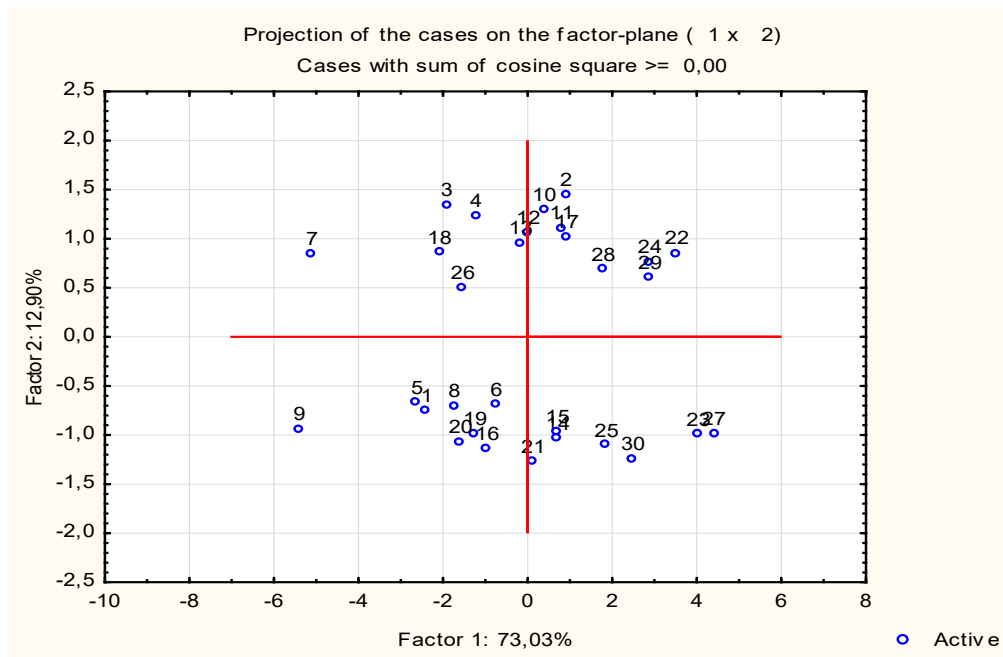


Figure 1. Simulated data

We will describe the arrangement of modeled objects (recall that each object is an employee with a confident set of features), which we will perform using the *Classification Learner* module from the *MatLab 2018b* program. This factor operates in manual means, when the techniques are tested, and automated, when the program looks at all probable answers to a given problem and indicates the best results. Once again, we underline that machine learning works on the basis of the classification of new notes which gives a more visual representation.

Figure 2 demonstrates classification outcomes obtained by means of the methods described above.

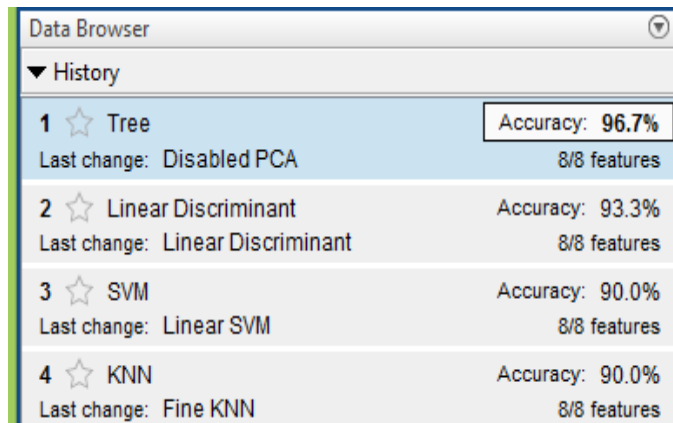


Figure 2. Results of algorithm schemes

Figure 2 shows that among the four indicated methods, the method of tree decision shows the best accuracy. It should be emphasized that the authors of the work do not seek to form a final conclusion on the conduct of a personnel audit, but only demonstrate the possibility of using the methods of machine learning to solve such a problem.

Figure 3 depicts error matrix with *TPR* and *FNR* on the right side.

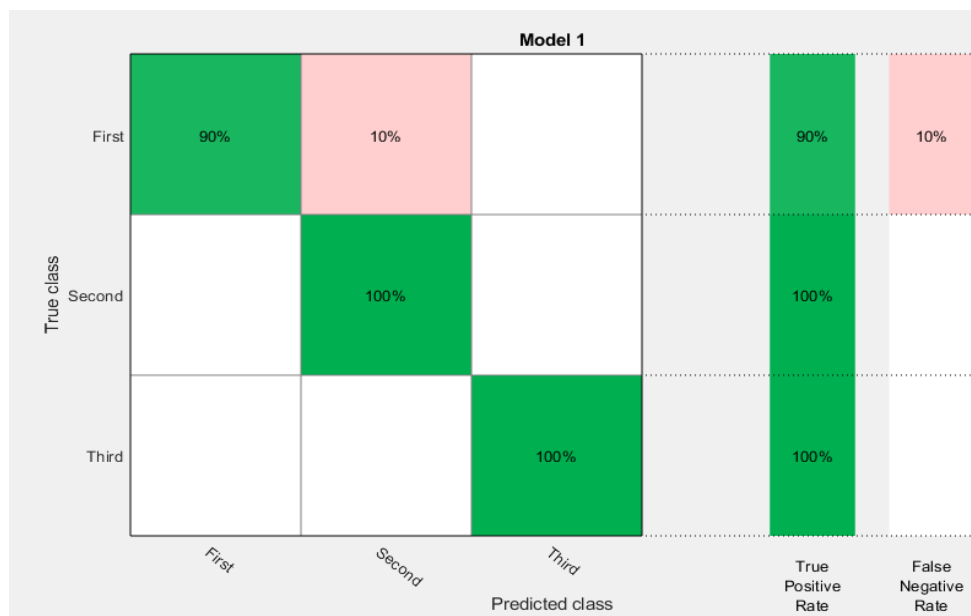


Figure 3. Matrix of errors

The TRP and FNR show data for every class. The upper stroke for the first class contains the values 90% and 10%. For this class 90% are correct, and 10% are erroneous ones. Objects of the 2nd and 3rd classes are separated without error: for them the values TPR = 100%.

To predict the class of a new observation, after creating the model, it is necessary to export it to the workspace, where the *trainModel* structure is created, which is used to evaluate the class with new information. We will also show a sample vector $au2 = [1; 28.0; 12.7; 9.7; 12.3; 10.3; 11.4; 12.6]$, which the model refers to the first class:

```
yfit = trainedModel.predictFcn(au2)
```

```
yfit =
```

```
categorical
```

```
First.
```

7. Conclusion

The results of the study on the choice of the method of decisions for the personnel audit, which were received by “manual” method of selection. They can be prolonged when switching to the programmed choice of classifiers. Then for such a task, it is advisable to consider the construction of a generalized audit assessment using fuzzy logic or to analyze the features that affect the efficiency of employees.

References

- Alpaydin, E. (2010). *Introduction to machine learning*, 2, Adaptive Computation and Machine Learning. Massachusetts Institute of Technology.
- Azad, M., & Moshkov, M. (2017). Multi-stage optimization of decision and inhibitory trees for decision tables with many-valued decision. *European Journal of Operational Research*, 263(3), 910-921.
- Coppin, B. (2004). *Artificial intelligence illuminated*. Jones & Bartlett Learning.
- Ezghari, S., Zahi, A., & Zenkouar, K. (2017). A new nearest neighbor classification method based on fuzzy set theory and aggregation operators. *Expert Systems with Applications*, 80, 58-74.
- Gao, W., Alsarraf, J., Moayedi, H., Shahsavari, A., & Nguyen, H. (2019). Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. *Applied Soft Computing*, 84, 105748.
- Kim, K., & Hong, J. A. (2017). Hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis. *Pattern Recognition Letters*, 98, 39-45.
- Personnel-audit. (2021). In: *FortySeven Software Professionals.com* [online]. <https://fortyseven47.com/services/personell-audit/>
- Personnel Audit. (2021). In: *ManagementMania.com* [online]. *Wilmington (DE) 2011-2021, 07/22/2015* <https://managementmania.com/en/personnel-audit>
- Principal Components Analysis. (2021). In: *www. NCSS Statistical Software.com* [online]. <https://www.ncss.com/>
- Ramsundar, B., & Zadeh, R. (2018). *TensorFlow for Deep Learning*. O'Reilly, Beijing-Boston.
- Vapnik, V. (1998). *Statistical Learning Theory*. J. Wiley.
- Yashkova, N. V., & Timofeeva, L. V. (2019). HR audit in the personnel management system. *Basic Research*, 2, 55-59.
- Zhang, X., Li, Y., & Zhang, X. (2017). KRNN: k Rare-class Nearest Neighbour classification. *Pattern Recognition*, 62, 33-44.