Computational supporting decision-making in self-government bodies using machine learning

Kh J Rakhimboev¹, M A Ismailov², H Kh Khusinov¹, Z A Hakimov¹ and O K Khujaev¹

¹Urgench branch of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, 110, al-Khwarizmi str., 220100, Urgench, Uzbekistan.
²Research and innovation centre of information and communication technologies at Tashkent University of Information Technologies, 17a, array Bozuv-2, Tashkent, Uzbekistan

E-mail: hikmatbek75@yandex.com

Abstract. The article discusses the problem of building a model and algorithm for supporting decision-making in self-government bodies by machine learning. As a method of machine learning, a multiple linear regression method is selected for processing the training sample. In the training sample, independent data consists of parametric assessments in the numerical form of self-government organs in three areas of activity, as, education, social environment and crime. A dependent parameter consists of generalized expert assessments of self-government bodies, also in numerical form. A model and an algorithm for the decision support process for making decisions using the multiple linear regression method is built. Based on the built model and the proposed algorithm, the ratios of the function coefficients to support the decision-making are identified. With this model, a generalized expert assessment in a numerical form is determined for the new self-government body, which is interpreted as a proposed solution to improve the state of the object.

1. Introduction
The adoption of management decisions is today an urgent topic, since for the effective and smooth operation of an organization, it is necessary to make management decisions correctly and in a timely manner. Nowadays, decision-makers are increasingly based on the analysis of data in a specific situation and are moving away from making decisions based on intuition. The managerial decision is concretized in the form of a program; it specifies a list of activities, methods of their implementation, terms and boundaries of actions, a range of performers, as well as planned results and criteria for their assessment.

While improving the quality of life of the population and for the development of the region, as well as self-government bodies, leaders should pay special attention to the economic and social state of the region [1]. The state of development of the region is assessed by the indicators of the corresponding field of activity.

2. Formulation of the problem
In this article, the object of research is self-government bodies [2], and the subject is decision support based on machine learning by linear regression. In this case, the main task is as follows: on the basis of
machine learning, numerical assessments of the state of development in the areas of activity of the self-government body and the corresponding decisions of experts expressed in numbers, support for decision-making for the new self-government body.

3. The structure of the training set
For machine learning of the above initial data, the linear regression method was chosen [3-5]. In a regression analysis problem, there is always a training set consisting of input parameters and responses, as well as an initial parametric model.

In the machine learning method using linear regression [6], used to solve the problem, the input data, that is, the trained set will consist of parametric estimates calculated on the basis of statistical indicators of the state of self-governments spheres of activity and expert assessments [7, 8].

The parametric assessment was carried out on the basis of relevant statistical data, that is, indicators by areas of activity, such as educational activities, the state of the social environment and crime in government. When calculating the parametric estimate, three methods were used to determine the weights of statistical indicators. As a result, the results of a parametric assessment of educational activity, social environment and crime were obtained using three different methods for determining the weights of indicators, in total, three parametric assessments were obtained for three areas of activity of one object [7]. The expert assessment for all three parametric assessments is the same. An example of parametric estimates for one self-government is given below (table 1):

| N   | Method for determining weights | Educational activities | Social environment | Crime   | Expert assessments |
|-----|--------------------------------|-----------------------|-------------------|---------|-------------------|
| 1.  | Summation (without weights)    | 0.392                 | 0.041             | -1.201  | 79                |
| 2.  | Rankings                       | -0.268                | 0.003             | -0.068  | 79                |
| 3.  | Attribution of points          | 14.482                | 0.003             | -0.078  | 79                |

Experts' estimates obtained in a percentage scale are used as dependent initial data. At the same time, expert assessments are interpreted as conclusions, decisions and proposals for solutions that are necessary to increase the efficiency and state of the areas of activity and lead to positive changes in the relevant areas of activity.

4. Solutions
Here, on the basis of a machine learning algorithm, adaptive functions are determined for bringing generalized parametric estimates of objects to expert estimates for the corresponding objects. The revealed adaptability function makes it possible to determine the expert assessment of a new object based on its parametric estimates.

The following figure shows the IDEF0 functional model for decision support based on machine learning linear regression [9].

From the IDEF0 model in figure 1 it can be seen that in the decision support process, the incoming information has a vector or matrix form and consists only of numerical data.

In this case, the incoming information is information in the form of a matrix for solving the problem of decision support. These are parametric assessments of the field of activity and expert assessments.

As a method, a method is used to determine the adaptability function using multiple linear regression, which affects the model.

Errors and approximations can be singled out as criteria influencing the model, as well as data output from the model of mining results.
Machine learning based on linear regression. Linear regression is a regression model of the dependence of one (explained or dependent) variable \( y \) on another or several other variables (factors, regressors, independent variables) \( x \) with a linear dependence function.

With paired regression, the model looks like this.

\[
y = f(x, b) + \epsilon, \quad E(\epsilon) = 0,
\]

where \( b \) - model parameters, \( \epsilon \) - random model error; is called linear regression if the regression function \( f(x, b) \) has the form:

\[
f(x, b) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k,
\]

where \( b_j \) are parameters (coefficients) of regression, \( x_j \) are regressors (factors of the model), \( k \) is the number of factors in the model [4].

In simple terms, linear regression is a search for a function, rather, finding the values of a function for a line that almost matches a given data so that future values can be predicted.

The linear regression coefficients show the rate of change of the dependent variable for a given factor, with the remaining factors fixed (in the linear model, this rate is constant):

\[
\forall j \quad b_j = \frac{\partial f}{\partial x_j} \quad \text{const}
\]

The parameter \( b_0 \), at which there are no factors, is often called a constant. Formally, this is the value of the function when all factors are zero. For analytical purposes, it is convenient to assume that a constant is a parameter with a "factor" equal to 1 (or another arbitrary constant, therefore this "factor" is also called a constant). In this case, if we renumber the factors and parameters of the original model taking this into account (leaving the designation for the total number of factors - \( k \)), then the linear regression function can be written in the following form, which does not formally contain a constant:

\[
f(x, b) = b_1x_1 + b_2x_2 + \ldots + b_kx_k = \sum_{j=1}^{k} x^T b_j
\]
where \( x^T = (x_1, x_2, ..., x_k) \) is a vector of regressors, \( b = (b_1, b_2, ..., b_k)^T \) is a column vector of parameters (coefficients).

The linear model can be with or without a constant. Then, in this representation, the first factor is either equal to one or is an ordinary factor, respectively.

Pairwise and multiple regression. In the special case, when the factor is the only one (without taking into account the constant), we speak of pairwise or simplest linear regression:

\[
y_i = a + bx_i + \varepsilon_i \tag{5}
\]

When the number of factors (excluding the constant) is more than 1, then they speak of multiple regression:

\[
Y = b_0 + b_1x_{i1} + ... + b_jx_{ij} + ... + b_kx_{ik} + \varepsilon_i \tag{6}
\]

Matrix representation of parametric assessments by areas of activity of the object. We have a sample of the number of \( n \) parametric estimates \( x_i \), which are obtained by three methods of parametric assessment and expert estimates (proposals and solutions) \( y_j \) for objects. Let us denote \( t \) - the number of the object in the sample. Then \( y_t \) — is the value of the \( y \) variable in the \( t \)-th object, \( x_{tj} \) — is the value of the \( j \)-th factor in the \( t \)-th object. Accordingly, \( X^T = (x_{1i}, x_{2i}, ..., x_{ki}) \) — is the vector of regressors in the \( t \)-th object. Then a linear regression relationship takes place in each observation:

\[
y_t = b_1x_{t1} + b_2x_{t2} + ... + b_kx_{tk} = \sum_{j=1}^{k} b_jx_{tj} = x^T_t b + \varepsilon_t, \quad \mathbb{E}(\varepsilon_t) = 0, \quad t = 1..n \tag{7}
\]

Introduce the notation:

\[
y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{— vector of expert judgment of the dependent variable } y; \tag{8}
\]

\[
X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad \text{— matrix of factors;} \tag{9}
\]

\[
\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \text{— vector of random error} \tag{10}
\]

Then the linear regression model can be represented in matrix form:

\[
y = Xb + \varepsilon \tag{11}
\]

5. Decision support algorithm based on machine learning with linear regression

The step-by-step view of the algorithm developed to solve the problem of linear regression learning of parametric estimates and expert estimates based on these formulas is as follows:

Stage 1: the calculated parametric estimates of the spheres of activity of each object are loaded into the array \( X[i,j] \), that is an independent factor matrix is formed, corresponding to formula (8);

Stage 2: dependent factors, consisting of expert assessments, form the vector \( y \) given in formula (7);

Stage 3: the least squares method calculates the values of the coefficients \( b_0, b_1, ..., b_k \) in formula (6);

Stage 4: the values of \( rt \) are calculated in the formula (6);

Stage 5: for the new object, a parametric assessment of the corresponding areas of activity is calculated;

Stage 6: based on the calculated values of \( b_0, b_1, ..., b_k \) and \( rt \), the corresponding value of the expert assessment is determined by the formula (11), which is proposed as a solution to the problem and improving the state for the current object.
6. Results
As a result of the calculation of the above algorithm, the values of the coefficients of the function of the proposed model are determined by the formula (2). For each method of determining the weight coefficients of indicators for parametric estimation [7], the function of the model, according to the results of the calculation given in tables 2-4, is smoothed as follows:

\[
f(x,b) = -0.469 - 0.1x_1 - 0.662x_2 + 0.523x_3 \quad (12)
\]
\[
f(x,b) = -0.103 + 0.27x_1 - 0.422x_2 + 0.362x_3 \quad (13)
\]
\[
f(x,b) = -0.167 + 0.281x_1 - 0.517x_2 + 0.375x_3 \quad (14)
\]

Table 2. Results without taking into account the weights of indicators.

| N  | Variables                  | Coefficient | Std. Err. | t-value | P>|t| |
|----|----------------------------|-------------|-----------|---------|-----|
| 1  | Evaluation of education   | \(-b_1 = 0.1\) | 0.051     | -1.972  | 0.049 |
| 2  | Social environment assessment | \(-b_2 = -0.662\) | 0.096     | -6.881  | 0     |
| 3  | Evaluation of crime       | \(b_3 = 0.523\) | 0.105     | 4.963   | 0     |
| 4  | The random error of the model | \(-b_0 = -0.469\) | 0.096     | -4.879  | 0     |

Table 3. The results when calculating the weights of the ranking indicators.

| N  | Variables                  | Coefficient | Std. Err. | t-value | P>|t| |
|----|----------------------------|-------------|-----------|---------|-----|
| 1  | Evaluation of education   | \(b_1 = 0.027\) | 0.042     | 0.65    | 0.516 |
| 2  | Social environment assessment | \(-b_2 = -0.422\) | 0.065     | -6.508  | 0     |
| 3  | Evaluation of crime       | \(b_3 = 0.362\) | 0.09      | 4.041   | 0     |
| 4  | The random error of the model | \(-b_0 = -0.103\) | 0.066     | -1.56   | 0.12  |

Table 4. Results when calculating the weights of indicators by attributing of points.

| N  | Variables                  | Coefficient | Std. Err. | t-value | P>|t| |
|----|----------------------------|-------------|-----------|---------|-----|
| 1  | Evaluation of education   | \(b_1 = 0.281\) | 0.044     | 6.375   | 0     |
| 2  | Social environment assessment | \(-b_2 = -0.517\) | 0.067     | -7.769  | 0     |
| 3  | Evaluation of crime       | \(b_3 = 0.375\) | 0.076     | 4.959   | 0     |
| 4  | The random error of the model | \(-b_0 = -0.167\) | 0.055     | -3.025  | 0.003 |

7. Conclusions
As a result of the study, a model of decision support in self-government bodies was developed based on a parametric assessment of the scope of the object and expert assessments. The model and algorithm of the decision support process using the method of multiple linear regression are built. On the basis of the constructed model and the proposed algorithm, the coefficients of the function for decision support are revealed. Using this model, a generalized expert assessment in numerical form is determined for a new self-government body, which is interpreted as a proposed solution to improve the state of the object.

References
[1] Demidov A A and Zakharov Yu N 2012 Information and analytical systems for decision-making support in state and local government bodies Fundamentals of design and implementation (Saint-Petersburg: NRU ITMO)
[2] Golovina A N 2018 Technology for making and implementing management decisions in local government bodies Ways to eliminate deficiencies Scientific and practical electronic journal Alley of Science 1(17)
[3] Voronina V V 2017 Theory and practice of machine learning Textbook (Ulyanovsk: UISTU)
[4] Demidenko E Z 1981 Linear and Nonlinear Regression (Moscow: Finance and Statistics)
[5] Barsegyan A A, Kupriyanov M S, Xolod I I, Tess M D and Elizarov S I 2009 Analysis of data and processes (Petersburg: BHV-Petersburg)

[6] Kamilov M M, Khujayev O K, Egamberdiyev N A and Chorshanbiyev A X 2019 The method of applying the algorithm of calculating grades for finding similar diagnostics in medical information systems International Journal of Innovative Technology and Exploring Engineering 8(6) 722-4

[7] Rakhimboev K J, Ismailov M A and Khalmuratov O U 2020 Parametric method for evaluation the state of complex organizational objects IOP Conf. Ser.: Mater. Sci. Eng. 919(5) 052043

[8] Rakhimboev K J and Ismailov M A 2020 Building a model for parametric assessment of the state of the control object and its components ACTA TTPU 219-33

[9] State standard of the Republic of Uzbekistan ISO/IEC 27035-2: 2016