Journal of Theoretical and Applied Information Technology
31st July 2021. Vol.99. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DATA MINING IN THE PROFESSIONAL EDUCATIONAL ENVIRONMENT USING MACHINE LEARNING METHODS

[1]AINAGUL AMANGELDINOVNA KABDIROVA, [2]VALENTINA PETROVNA KULIKOVA, [3]VLADIMIR PAVLOVICH KULIKOV

[1,2,3]North Kazakhstan University named after Manash Kozybayev, Petropavlovsk, Kazakhstan

E-mail: a.kabdirova@mail.ru

ABSTRACT

The article deals with the problems of building an intelligent system for organizing course training using machine learning methods. The main idea of the present study is to solve the problem of optimal layout and the presentation rate of the material for the listener to learn the courses by automating the methods of data mining. Machine learning is used to collect and initially process information about the trainees for its structuring, identification of characteristic features, generalization, and sorting. Proper evaluation of this information helps to build a more correct and adequate model of the learning process for subsequent analysis and planning. The authors propose automated mechanisms for organizing a personalized environment in the context of managing the development of professional self-actualization of a specialist based on training models such as the "static" case, loss of qualifications, lost profits, and a dynamic model. The article describes the possibilities of using multidimensional regression analysis in predicting the behavior of the system under study for making managerial decisions and possible risks. The results of the conducted study will be useful for researchers in the field of software design and mathematical simulation, analysts, as well as vocational education teachers – for integrating information data about listeners.

Keywords: *Advanced training, Management, Automation, Data analysis, Dynamic model*.

1. INTRODUCTION

Although machine learning methods are already used in many fields, this technology has not yet been widely used in education. However, the relevance of this topic is undeniable, since the importance of the development of artificial intelligence technologies for the sciences involving the study of the management of various facilities as well as groups of people increases. It is also very important to implement machine learning for the collection and initial processing of information about the trainees for its structuring, identification of characteristic features, generalization, and sorting. Proper evaluation of this information is expected to contribute to the construction of a more correct and adequate explanatory model of the learning process for subsequent analysis, planning, and management decision-making [1]. For example, at the current stage of economic development in Kazakhstan, there is no single viewpoint on minimizing any personnel risks through the system of additional vocational education. In particular, this concerns the organization of the analysis and evaluation of the processes of changing the qualifications of employees to improve the efficiency of the economy.

It is very important to analyze and evaluate risks, which is due to the large uncertainty of the impact (usually negative) on labor potential and financing. The proficiency enhancement is a complex structured process that requires analysis of the interaction of risk-forming factors, monitoring the quality of course training, and evaluating the efficiency of training courses. Therefore, the applicability of classification and clustering processes in terms of building an automated intelligent system is justified and comprehended (Figure 1).

In the course of proficiency enhancement, *the subject context of professional activity remains dominant*, namely, knowledge, skills, abilities (KSA), experience in using KSA, and responsibility for the consequences of using specific KSA in the decision-making. *The social context is excluded,* which opens up the ways and opportunities for the specialist to join the team, form the ability of social interaction and communication, joint decision-making, personal and corporative responsibility for any activity [2].

Journal of Theoretical and Applied Information Technology
31st July 2021. Vol.99. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 1: Components of the Intellectual Educational System of Course Preparation Organization

The idea of the presented system is to solve the problem of optimal structure and rat1e of presenting material to the listener when learning the courses.

At that, the model can contain the characteristics of a specialist, for example, such as *strong*, *good*, *intermediate*, and *weak*, as well as databases on knowledge of a particular subject area; clustering of groups based on input and output data; and analysis of risk factors.

Thus, the complexity and variety of data mining methods will require specialized end-user tools suitable for solving certain typical information analysis tasks in specific areas of education.

In turn, the results of dynamic data analysis described in the work, based on the use of mathematical models and risk assessment methods to intensify the process, can be useful to researchers as a step-by-step information support toolkit when making management decisions.

2. MATERIALS AND METHODS

The possibility to use well-known methods of mathematical statistics and machine learning to solve problems of this kind has opened up new opportunities for analysts, researchers, as well as those who make decisions, i.e. managers and company executives.

There are several key methods of machine learning. The first method is learning with a teacher. With this method of training, a pair of "situation and required solutions" is given for each case. In the concerned case, the main goal of such training is to find the dependence of the decision made in a given situation and create an appropriate algorithm that can describe the situation as input data and predict the proper solution. A huge set of training data can be entered into the computer, and the parameters can change until the output results of the course do not match the expected outcomes. Learning without a teacher is one of the methods of machine learning, in which the system under test spontaneously performs a task without the intervention of the experimenter. A partial training option is also possible when the training is built using a small amount of unmarked data. In this case, the system makes a decision based on the marked data and inherent computational variables. Many machine learning researchers have found that unmarked data, when used in combination with a small amount of marked data, can significantly improve the accuracy of training [3].

To improve the training accuracy and the reliability of the results, a feature vector of the objects is used. It is assumed that each object of analysis is described by a certain set of features: nominal, binary, ordinal, and numeric.

All features are divided into a subset of features known for all objects without exception, and a subset of features known only for the objects of the learning sample. For the objects of the test and working samples, they need to be interpreted using the trained model.

Let X be a set of objects, while Y be a set of answers, $y : X \to Y$ is an unknown dependence. Then the machine learning problem in the most general form is formulated as follows. Given parameters are training sample $\{x_1,...x_l\} \subset X$, and known answers $y_i = y(x_i)$, $i = \overline{1,l}$, where l is the number of objects in the training sample. It is necessary to find an algorithm $a : X \to Y$, or a decision function, approximating y on the entire set of X.

As noted above, the most common way to specify objects is their feature description: $f_j = X \to D_j$ where f_j $j = \overline{1,n}$ are features of the objects, n is the number of the object features, D_j is the range of attribute instances; at that, $D_j = \{0,1\}$, if f_j is a binary feature; $|D_j| \langle \infty$ if f_j is a nominal feature; $|D_j| \langle \infty$ and we enter the ordering comparison for D_j if f_j is an ordinal feature; and $D_j = R$, if f_j is a quantitative feature. A vector $(f_1(x),..., f_n(x))$ is a feature description of an object x. The task of learning from examples deals with the objects-features matrix:

$$F = \|f_i(x_i)\|_{l\times n} = \begin{pmatrix} f_1(x_1) & \ldots & f_n(x_1) \\ \ldots & \ldots & \ldots \\ f_1(x_l) & \ldots & f_n(x_l) \end{pmatrix}.$$

1. The model training stage is that the training method $\mu : (X \times Y)^l \to A$ based on the sample $X^l = (x_i, y_i)_{i=1}^l$ builds an algorithm $a = \mu(X^l)$:

$$\left| \begin{pmatrix} f_1(x_1) & \ldots & f_n(x_1) \\ \ldots & \ldots & \ldots \\ f_1(x_l) & \ldots & f_n(x_l) \end{pmatrix} \to \begin{pmatrix} y_1 \\ \ldots \\ y_l \end{pmatrix} \right| \xrightarrow{\mu} a.$$

2. The model application stage is that the algorithm a for new objects $x_1',...x_k'$ gives answers $a(x_i')$:

$$\begin{pmatrix} f_1(x_1') & \ldots & f_n(x_1') \\ \ldots & \ldots & \ldots \\ f_1(x_k') & \ldots & f_n(x_k') \end{pmatrix} \xrightarrow{a} \begin{pmatrix} a(x_1') \\ \ldots \\ a(x_k') \end{pmatrix}.$$

To evaluate the quality of the algorithm a, i.e. the proximity of a to y, the quality functionals are used. For the point estimation, the loss function is used, $\alpha(a,x)$ is the error of the algorithm $a \in A$ on the object $x \in X$, for the integral estimation – the empirical risk is used

$$Q(a, X^l) = \frac{1}{l}\sum_{i=1}^{l} \alpha(a, x_i),$$

which serves the quality functional of the algorithm a on X^l [4].

Thus, the goal of machine learning can be formulated as minimizing empirical risk:

$$\mu(X^l) = \arg\min_{a \in A} Q(a, X^l).$$

The selected tasks in the system of additional vocational education can be implemented using appropriate machine learning algorithms based on individual indicators and characteristics.

To solve one of the machine learning problems, let us use discriminant data mining. Discriminant data mining, based on the use of past experience, allows predicting the behavior of new observations (objects) and determining their belonging to a certain group based on the principle of maximum similarity of the characteristics of new observations with the features of the training sample. The advantage of discriminant data mining in comparison with other classification methods is that when constructing the discriminant function, the variables that distinguish the groups in the best way are determined automatically [5, 6].

In n-dimensional space, we are talking about finding a linear combination of the form

$$F(x) = \sum_{n=1}^{n} a_n x_n$$

where $x_1 \ldots x_n$ are discriminant variables, $a_0 \ldots a_n$ are weights of discriminant variables, n is the number of variables.

The discriminatory analysis procedure consists of two main parts:

- determining the most informative characteristics (discriminant variables) of the groups from the training sample, and constructing the discriminant function on this basis;
- classifying new objects (observations) using the resulting model, by substituting the values of the variables of each object in the discriminant function.

Taking into account the requirements for discriminant variables that distinguish one class from another, 75 school teachers with different individual professional characteristics involved in the upgrade training course, were selected for the present study.

1) Binary features: gender, availability of e-mail, territorial affiliation (city/village), winner of national and international competitions of professional skills, author/co-author of textbooks, compiler of methodological manuals, and recommendations.

2) Ordinal features: education (secondary, higher, postgraduate), qualification category (no category, second, first, higher category).

3) Nominal features: the status of the educational organization (lyceum/gymnasium, general education, primary, low-grade school).

4) Quantitative features: age, teaching experience, the results of the entrance subject testing, self-assessment based on personal data, the degree of proficiency in educational technologies.

Five parameters that meet all the requirements for discriminant variables (a fragment of data is presented in Table 1) were considered as discriminating variables. The grouping categorical variable is the Partitioning Class that characterizes the group of distributed training depending on the dynamics of professional activity indicators: A(3) – advanced level, B(2) – intermediate level, and C(1) – basic/elementary level.

Table 1: Fragment of the Source Data

№	Name	Subject testing results (%)	Teaching experience (years)	Qualification category 0-no category 1-second 2-first 3-higher	Teacher's self-assessment (average score)	Technology proficiency (%)	Partition class
1	Abenova	46	1	0	4	40	
2	Abilkasova	89	40	2	4,25	80	3
3	Abilova	54	2	0	4,25	45	
4	Abilkina	61	8	1	4,5	65	
5	Alibekova	78	25	3	4	85	3
6	Alzhanova	73	24	2	4,75	75	2
7	Alkina	68	5	1	4	35	2
8	Baizhigitova	46	1	0	3,75	45	1

The training sample is represented by objects from each training group in the following proportions: 14 listeners belong to the basic level, 14 listeners – to the intermediate level, and 12 listeners – to the advanced level. The distribution of listeners for the training sample (based on the criterion of the initial qualification level) was carried out by an expert method. Listeners with a level of subject knowledge above 76% were included in group A(3); listeners with a level of less than 59% were included in group C(1), the rest – in group B(2).

The task was set to classify listeners who were not included in the training sample, namely:

- defining the criterion by which observations can be classified into the category of the Partitioning Class depending on the values of the parameters {Var2=Results of subject testing, Var3=Teaching experience, Var4=Qualification category, Var5=Self-assessment, Var6=Technology proficiency};
- defining the Partitioning Class category for the new observation.

The initial data for the training sample is presented as a matrix for each group:

$$X_i = \begin{pmatrix} x_{11} & \dots & x_{n1} \\ \dots & \dots & \dots \\ x_{1m} & \dots & x_{nm} \end{pmatrix}$$

where, i is the designation of the group (i={A, B, C}); n is the number of variables; m is the number of observations; x_{nm} is the value of the n-th variable of the m-th observation.

Next, the classification functions for each group are constructed. The classification function is a linear combination of features that are designed to determine which group the observations are most likely to belong to.

$$F_A(x) = c_A + b_{A1} * x_1 + b_{A2} * x_2 + \dots b_{An} * x_n$$
$$F_B(x) = c_B + b_{B1} * x_1 + b_{B2} * x_2 + \dots b_{Bn} * x_n$$
$$F_C(x) = c_C + b_{C1} * x_1 + b_{C2} * x_2 + \dots b_{Cn} * x_n$$

where b is the coefficients of the discriminant function or the weights of the corresponding variables for each class, c is the regression constants; $x_1 \dots x_n$ are the values of features (independent variables, or predictors).

A probabilistic classification method is used, in which an object should be assigned to the class (general population) within which it looks plausible, i.e. according to the criterion of maximizing the classification indicator $F(x)$.

This classification will allow developing scenarios for taking measures to minimize negative risks, as a response to changes in the system of organizing course training in connection with the emergence of risk-forming factors.

The scheme for building a risk management model must necessarily include the following.

1. Defining indicators and parameters, for example:

– From n to N listeners are invited (addressed).

– They form groups in the number from k to K.

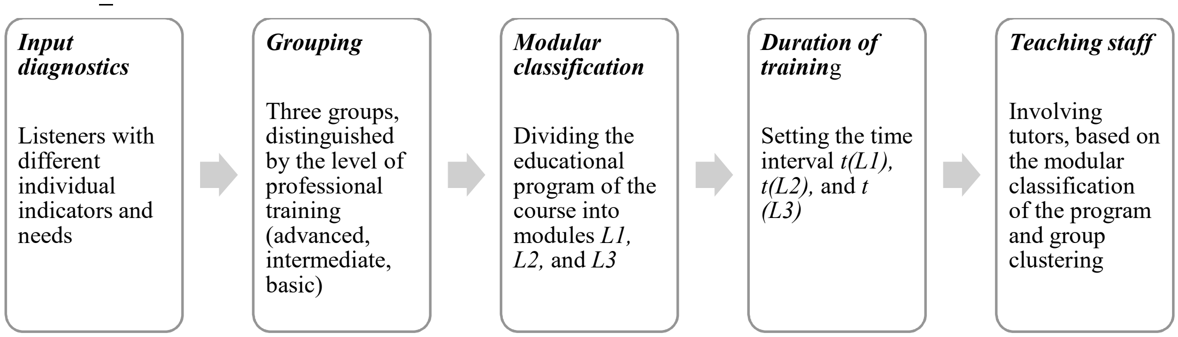
– Groups vary in level of training from u to U.

– The course content is divided into modules (stages) in the number from l to L.

– The duration of the stages is within the range from t to T.

– Tutors in the number from m to M are involved.

Educational modules are equipped with tests (test papers) in the number from z to Z, including input and output control.



ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. Forming a series of questions to develop scenarios like "what if?", for example:

– The number of applying listeners is more than the established maximum or less than minimum;

– The number of groups is greater or smaller than expected;

– Groups are below or above the intended course level;

– It is needed to recompose/split/expand the modules;

– It is necessary to exclude modules, reorder them;

– The training stage(s) will be greatly shortened or stretched over time;

– A force majeure situation in terms of the required number of tutors may occur.

To substantiating further the effectiveness of tendentious forecasting, the authors use the extensive capabilities of regression and correlation analysis:

- evaluating the correctness of the developed regression model (specification);

- assessing the quality of the empirical regression equation (i.e., the degree of compliance with actual statistical data, which will allow assessing the reliability of the identified pattern), i.e. checking the properties of the data, whose feasibility was assumed when evaluating the equation; checking the statistical significance of the regression equation coefficients; checking the overall quality of the regression equation, etc.

Regression-correlation analysis is performed for a limited sample, therefore the parameters of the regression equation (regression and correlation indicators), the correlation coefficient, and the determination coefficient can be impacted by random factors. To check the nature of the impact of these indicators on the entire general sample, it is necessary to check the adequacy of the constructed statistical models.

When analyzing the adequacy of the regression equation (model) to the process under study, the following options are possible:

• the model based on the Fischer F-test is generally adequate, and all regression coefficients are significant (the model can be used for making decisions and implementing forecasts);

• the model based on Fischer's F-criterion is adequate, but some of its coefficients are insignificant (the model is suitable for making certain decisions, but it is not suitable for implementing forecasts);

• the model based on the F-criterion is adequate, but all the regression coefficients are insignificant (the model is suitable neither for providing support for decision-making nor implementing forecasts).

To check the significance (quality) of the regression equation means to determine whether the mathematical model expressing the dependence between the variables corresponds to the experimental data; as well as whether the explicative variables included in the equation are sufficient to describe the dependent variable. To determine the quality of the model, the average approximation error is determined for each observation from the relative deviations [4, 7].

It is possible to assume the uniform nature of the dependence of all Y indicators on x_i only if the sample population is sufficiently homogeneous. In practice, the presence of clearly expressed different groups of course training does not allow modeling the final dependence based on a single set of indicators.

Note that the Fisher discriminatory analysis procedure is computationally/interpretatively, similar to the multiple regression procedure:

- step-by-step discriminatory analysis is based on the use of a statistical significance level;

- the variables with the highest regression coefficients contribute the most to discrimination.

3. RESULTS

For the preliminary selection of model variables (distribution parameters in the classes assigned by experts), let us estimate the mathematical expectations of the features in the classes (Figure 2). Except to Var5=Self-assessment, all the average values of the indicators calculated by the first training sample of the Partitioning Class are visibly less than the corresponding average values of the indicators calculated by the second training sample, whose average values are less than the similar values of the third sample. Without testing the significance of discrepancies in the average values of the indicators, it can be assumed that the considered indicators of professional changes of listeners correspond to the expert level of the distribution.

The most discriminating variables are Var2=Subject test results, Var3=Teaching experience, and Var6=Technology proficiency. The variable Var4=Qualification category does not distinguish between listeners with an initial and intermediate level of training. Var5=Self-assessment does not distinguish well between all listeners, as well as between high-level and low-level listeners. The content and application of the questionnaire should be reviewed, first of all, to exclude bias in the self-assessment parameter, rather than the idea of excluding the parameter itself.

	Means (Spreadsheet1)					
Var7	Var2	Var3	Var5	Var4	Var6	Valid N
G_1:1	53,50000	4,42857	4,125000	0,357143	53,21429	14
G_2:2	70,14286	15,71429	4,089286	1,357143	64,64286	14
G_3:3	84,33334	23,83333	4,125000	1,833333	75,83334	12
All Grps	68,57500	14,20000	4,112500	1,150000	64,00000	40

Figure 2: Estimation of Mathematical Expectations of Features in Classes

The step-by-step discriminatory analysis is performed in two ways: step-by-step analysis with inclusion of variables, and step-by-step analysis with their exclusion. In the first method, at each step, all variables are viewed and those that make the greatest contribution to the difference between the samples are selected, while in the second method – all variables are first included in the model, and then at each step, variables that make a small contribution to the prediction are excluded (Figure 3).

	Discriminant Function Analysis Summary (Spreadsheet1) No. of vars in model: 5; Grouping: Var7 (3 grps) Wilks' Lambda: ,13124 approx. F (10,66)=11,619 p< ,0000					
N=40	Wilks' Lambda	Partial Lambda	F-remove (2,33)	p-value	Toler.	1-Toler. (R-Sqr.)
Var2	0,548069	0,239456	52,40629	0,000000	0,710049	0,289951
Var3	0,137263	0,956104	0,75753	0,476801	0,352019	0,647981
Var4	0,144532	0,908025	1,67131	0,203522	0,577225	0,422775
Var5	0,133104	0,985983	0,23457	0,792218	0,915769	0,084231
Var6	0,132801	0,988234	0,19645	0,822598	0,477023	0,522977

Figure 3: Results of Discriminatory Analysis

Wilks' lambda is a ratio of the measure of intra-group variation to the total measure of variation, which characterizes the degree of the model discrimination [7]. The closer its value is to zero, the better the statistical significance of the discrimination power. In the concerned case, the Wilks' lambda, in general, is equal to 0.13124, F=11.619 at a p-value <0.00001, that is, it can be stated that discrimination between groups is highly significant.

In terms of the independent contributions of each model variable to the overall discrimination between groups of listeners, the following can be noted.

The Wilks' lambda values in the first column show the values for the model where the corresponding variable is not used. From the results obtained, it can be seen that without the variable Var2=Subject test results (the Wilks' lambda is approximately 0.55), the quality of discrimination would be significantly worse. Next, the variables are ranked by significance as follows: Var4, Var3, Var5, and Var6.

Partial Lambda (statistics for the single contribution of the corresponding variable to discrimination) is the ratio of the Wilks' lambda after adding this *variable* to the Wilks' lambda before *adding* it [8]. In other words, the lower is the value of this indicator, the more valuable this variable is, i.e. the greater is its contribution to the overall discrimination. The value of Var2=Subject test results is confirmed by the value of the partial contribution: the Partial Wilks' lambda is approximately 0.24. The ranking of other variables has not changed: Var4, Var3, Var5, and Var6.

The F-include value is the criterion for including the variable in the model (the higher the value, the more significant). The P-value indicates the statistical significance of these variables (the lower, the more significant). Tolerance (calculated as 1-RSqr, where RSqr is multiple correlations with all other variables included in the model) is interpreted as a measure of the redundancy of a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

variable in the model (the lower the value, the less additional information is carried by the variable).

The results of the discriminatory analysis confirm the preliminary discriminating ability of the variable Var2=Results of the subject testing, which can be successfully included in the model. The discriminating ability of the other variables is doubtful; nevertheless, they remain in the model:

- in general, the particular p-values should not be taken as the decisive result;

- from meaningful model assumptions.

The linear Fischer discriminant functions (classification functions) for each group are as follows:

$$\textit{Advanced:}\ F_A(x) = -199{,}440 + 3{,}388 * x_1 - 1{,}128 * x_2 + 25{,}353 * x_3 + 1{,}442 * x_4 + 0{,}401 * x_5$$

$$\textit{Intermediate:}\ F_B(x) = -158{,}001 + 2{,}828 * x_1 - 1{,}021 * x_2 + 25{,}589 * x_3 + 0{,}632 * x_4 + 0{,}404 * x_5$$

$$\textit{Basic:}\ F_C(x) = -123{,}721 + 2{,}153 * x_1 - 0{,}908 * x_2 + 26{,}805 * x_3 - 1{,}097 * x_4 + 0{,}451 * x_5$$

where x_1 is the subject test results; x_2 is the teaching experience; x_3 is the qualification category; x_4 is the self-assessment of basic knowledge, and x_5 is the proficiency of educational technologies.

Thus, there are quite similar regressions (discriminant functions) by all parameters, except for the parameter x4=Var5=Self-assessment (the expectedly interpreted result: low-knowledge listeners tend to overestimate their knowledge, while high-knowledge listeners tend to underestimate). A statistically significant result illustrates *the irreducibility of the approach to a one-dimensional multivariate model*.

The procedure for constructing a classification matrix that allows obtaining fairly good discrimination (for example, a 100% correct distribution for at least one group) consists of several iterations that exclude specific objects with a mismatch from the samples. In the resulting matrix (Figure 4) one object from the third group was classified incorrectly using the obtained Fisher discriminant functions and fell into the second group). Despite the error in the assessment of one listener with subject knowledge of more than 76% (from group A(3)) by attributing to him the characteristics of the listener with the knowledge level of not more than 76%, but more than 59%, and relating him wrongly to group B(2), the total percentage of correct discrimination is quite high and equals to 97.5%.

Group	Classification Matrix (Spreadsheet1) Rows: Observed classifications Columns: Predicted classifications Percent Correct	G_1:1 p=,35000	G_2:2 p=,35000	G_3:3 p=,30000
G_1:1	100,0000	14	0	0
G_2:2	100,0000	0	14	0
G_3:3	91,6667	0	1	11
Total	97,5000	14	15	11

Figure 4: The Results of the Classification of the Training Sample Objects

The visual separation of the observational data is shown in Figure 5.

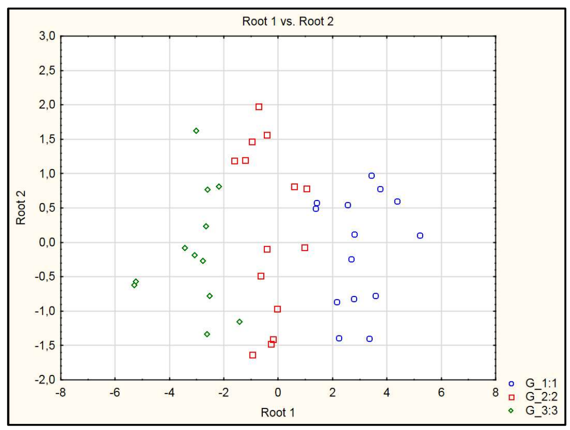


Figure 5: Scattering Diagram Based on Discriminant Functions

For prospective studies, a priori probabilities are used, which increases the classification accuracy. At this stage, the default setting is adopted – the probabilities of the course participant belonging to a certain study group are proportional to the size of the groups.

So, the Fischer discriminant functions are used as a diagnostic means of identification for each of the A, B, and C groups, according to which the discrimination/classification of the listeners involved in the construction of the corresponding functions was carried out quite successfully.

The Mahalanobis metric is used as a classification procedure that allows "predicting the behavior" of newly arrived listeners in terms of a dependent variable (according to the category of the Partitioning Class) based on the measurement of several independent variables/factors/indicators that

Journal of Theoretical and Applied Information Technology
31st July 2021. Vol.99. No 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

characterize them (Var2=Subject test results, Var3=Teaching experience, Var4=Qualification category, Var5=Self-assessment, Var6=Technology proficiency).

So, for the distribution of new listeners among significantly selected groups in the multidimensional space defined by the variables of the constructed discrimination model, the Mahalanobis distances from the listeners' features (new objects of observation) to the centroids of each of the classes were calculated. The new objects (35 listeners) will belong to the group the distance to which is the smallest (Figure 6).

Case	Squared Mahalanobis Distances from Group Centroi Incorrect classifications are marked with *			
	Observed Classif.	G_1:1 p=,35000	G_2:2 p=,35000	G_3:3 p=,30000
1	---	3,16308	23,37435	56,73753
2	G_3:3	45,56723	14,07467	4,67381
3	---	0,87995	11,60245	36,36489
4	---	3,87239	5,54125	22,33460
5	G_3:3	30,86400	7,20529	5,20555
6	G_2:2	16,16428	3,36894	8,68182
7	G_2:2	19,46756	6,97930	14,92736
8	G_1:1	3,11274	23,18469	56,45243
9	G_2:2	14,25463	4,63844	11,38080
10	G_3:3	70,30318	27,13698	7,68509

Figure 6: Mahalanobis Distances

Thus, the newly arrived listeners of advanced training courses with a reliability of 97.5% are classified (distributed) into groups of personalized training as follows: group A(3) advanced level of knowledge – 12 teachers; group B(2) intermediate level – 11 teachers; group C(1) basic level – 12 teachers.

The table of posterior probabilities gives similar results, taking into account the knowledge of the values of variables in the model. Even with the presence of incorrectly classified listeners, the classification accuracy is quite high.

Next, a model Y_{ij} (a system of equations) is involved which reflects the dependence of the training time to the *i*-th training module required for the assimilation of the material by the *j*-th subgroup of listeners, from the internal assessment of the subject knowledge x_{ij} and the volume of the modules of the educational program p_{ij} (important: *i* is determined by the course design structure by modules and runs through the range of numbers up to the current number of the module):

$$Y_{ij} = a^0 + a^1 x_{ij} + a^2 p_{ij} + \varepsilon_{ij} .$$

Here $\sum_i Y_{ij}$ in the *j*-th subgroup will be equal to the duration of the subgroup training.

If the modules in different subgroups may substantially coincide in terms of volume and meaning (certainly, the timings may not coincide), then it is possible that $p_{ij} = p_{km}$ (the additional balance dependencies may appear).

4. DISCUSSION

Integration, efficiency, and assessment of the system of retraining and advanced training of specialists in various fields are considered in the works of various scientists.

The monograph [9], devoted to the mathematical and computational foundations of artificial intelligence, presents the design of intelligent computing agents, the use of machine learning methods from the point of view of a multidimensional design space. In the article [3], the authors study the problem of big data and methods of storage, processing, and visualization in distributed database systems. The solution of classification and forecasting problems using graphical probabilistic models was studied in [7, 8]. In articles [10, 11], the area of research is focused on visualizing the research structure of professional competencies to identify their basis, development, and evolution over time, paying more attention to the technology and engineering cluster.

Thus, researchers, using mathematical modeling methods, have developed several tools for data mining, inference, and forecasting.

However, in the context of determining the level of assimilation and effectiveness of the learning process, the responsibility for making incorrect management decisions increases. Thus, the issues related to the assessment of risks that may arise in the process under consideration are actualized.

In the described study, we proposed a set of related machine learning methods, in particular, discriminatory analysis, allows getting answers to questions such as, for example:

- Is it possible to distinguish between the groups of listeners of proficiency enhancement (retraining) courses according to certain criteria? (in the formal terms the question concerns whether formulated categorical variable is grouping?)

- What are the parameters/factors/features that describe the structure of the set of proficiency enhancement courses that have the highest diagnostic ability? (in the formal terms the question

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

concerns discriminating variables that affect the discrimination of the groups most significantly?)

- How well do the selected parameters/factors allow distributing the listeners of the courses into groups of personalized training? (in formal terms, the question concerns how well is the discrimination/classification of the objects observed carried out?)

Fischer discriminant functions can be successfully used as a diagnostic tool for identifying listeners of advanced training courses according to certain input criteria. Based on the training samples, the listeners who were not included in the training samples were successfully re-classified, as well as the newly arrived listeners that were subject to grouping. The criterion for assigning to a particular group is the *maximum of classified value*. At that, similar results were obtained by classification based on:

- classification functions (according to the value of F(x) – evaluation of the category of the Partitioning Class);

- discriminant functions (according to statistical criteria employing a table with posterior probabilities).

The results of discriminatory analysis of advanced training course statistics are suitable for making some managerial decisions, for example:

- having information about a certain number of listeners with a knowledge level of less than, for example, 59% (group C), that may be resulted from several factors Var, one can build a function that allows determining which factor (factors) in particular caused the reason for being assigned to the entry-level group, and developing corrective (propaedeutic) measures before taking advanced training courses;

- it is possible to predict (determine the probability) the success of advanced training, if the training group is formed taking into account the performance level and, accordingly, to make a decision on the feasibility of conducting courses and/or including certain potential listeners in the training group;

- the discrepancy between the expert opinion on the parameters (characteristics) of the distributed learning group members and the statistical significance of the discriminating variables may initiate the reprocessing/modification of control and measurement materials.

The following advancements in machine learning are expected in the future:

- Machine learning methods should reasonably become an integral part of an automated data mining system for the high-quality organization of the learning process and professional self-actualization, taking into account the emerging risks (Figure 7).

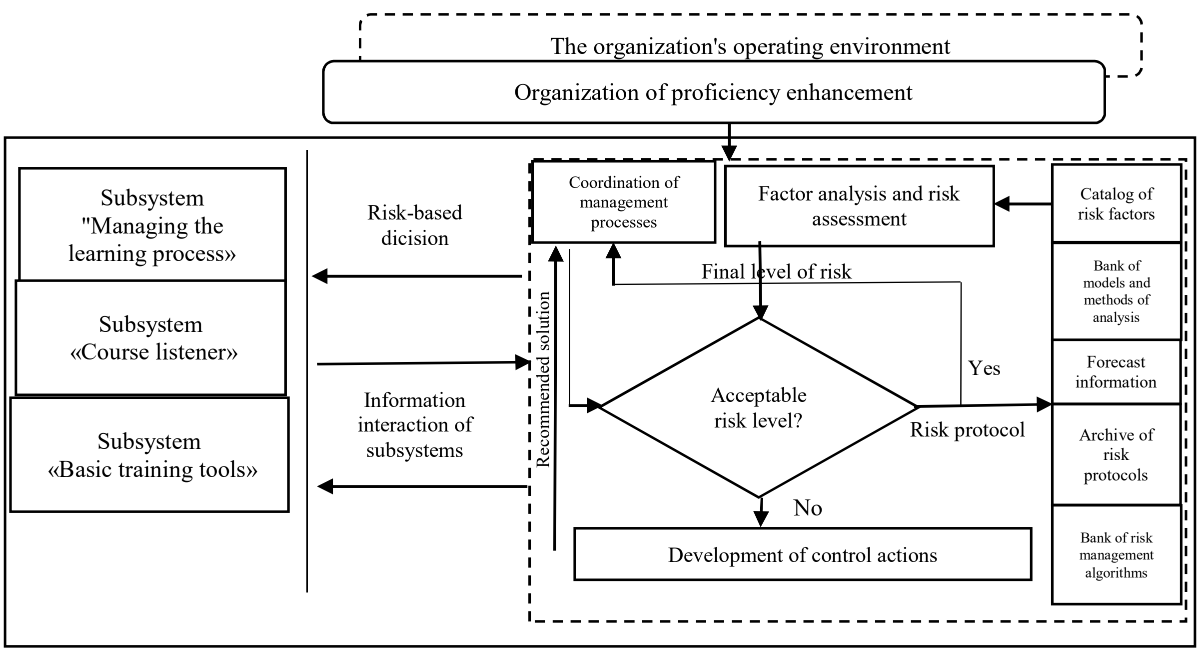


Figure 7: Scheme of the Risk Management Algorithm [12]

- For significantly different groups of course listeners, it is advisable to build individual, though interrelated models. For example, systems of simultaneous equations.

- When assessing the risks of violating the information integrity of the course, it is necessary to use additional formal criteria that allow determining the borders of indicators for which the same mechanisms are applicable (which does not mean that the regression model is reducible to a one-dimensional one, even if it is multifactorial), for example, entropy estimates and/or models based on Zipf-Pareto statistics (if the professional development system structure is justified as cenological).

The development of "what-if" scenarios will provide information support for making decisions concerning:

- possible increase in the number of listeners;
- the correct distribution of financial resources (the increase in costs depending on the number of tutors involved, the time interval, etc.);
- the completeness and depth of the educational program content in terms of taking into account the needs of regional education and individual educational requests of listeners;
- objectivity of testing and assessment materials;
- the effectiveness of the course, based on the results of the output diagnostics.

5. CONCLUSION

The classification methods proposed for information technology professionals were analyzed along with the importance of data mining, development of control systems, and support of computer technology in this area.

The proposed mathematical toolkit facilitates the structuring and organization of the process of optimal layout and the rate of material delivery for the learners to assimilate, which is very useful for analyzing the information integrity of the course and supporting managerial decision-making.

The controversial nature of the applicability of the described approach to the newly discovered areas of retraining should be noted. In particular, for the case of a pre-selected contingent of trainees who previously demonstrated a relatively indistinguishable level of basic training.

REFRENCES:

[1] N.Yu. Khafizova, N.V. Ekimova, and I.B. Bykova, “The growth of the level of professionalism of a teacher through informal professional development in the information and educational environment”, *Journal "Modern Problems of Science and Education"*, No. 5, 2017, p. 324.

[2] H. Beetham, and R. Sharp, *Rethinking of pedagogy for the digital age. Design in the education of the XXI century*. Almaty, Kazakhstan: Public Fund "Ulttyk audarma burosy", 2019, 352 p.

[3] Z.S. Seydametova, “Machine learning tasks and algorithms: probabilistic graphical models”, *Information and Computer Technologies in the Economy, Education and Social Sphere*, Vol. 1, No. 23, 2019, pp. 180-187.

[4] K. de Fine Licht, and J. de Fine Licht, “Artificial intelligence, transparency, and public decision-making”, *AI & SOCIETY*, Vol. 35, 2020, pp. 917-926. https://doi.org/10.1007/s00146-020-00960-w

[5] Zh. Kerimkhan, “Discriminatory analysis in the assessment of credit risks”, *Journal "Banks of Kazakhstan"*, No. 9, 2018, pp. 16-28.

[6] V.V. Kislitsina, “Discriminatory analysis as a tool for classifying business entities”, *Journal "Science and Education: New Time"*, Vol. 2, No. 19, 2017, pp. 50-58.

[7] J.K. Kangas, *Combinatorial Algorithms with Applications in Learning Graphical Models*. Series of Publications A. Helsinki, Finland: University of Helsinki, 2016, 66 p.

[8] S. Dorairaj, “Hidden Markov Models Simplified”, *Medium: medium taps into the brains of the world's most insightful writers, thinkers, and storytellers*. [Online]. Available: https://bit.ly/2HEryrQ

[9] D.L. Poole, and A.K. Mackworth, *Artificial Intelligence: Foundations of Computational Agents*. 2nd ed. New York, USA: Cambridge University Press, 2017, 820 p.

[10] *Amazon Machine Learning. Developer Guide*, AWS Documentation: user guides, developer guides, API references, tutorials, and more. [Online]. Available: https://docs.aws.amazon.com/machine-learning/latest/dg/index.html

[11] D. Guerrero, G. La Rosa, P. Lopez, A.L. Bayona, “Domain Analysis of the Research

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in Professional Competences, Technology and Engineering Cluster", *Procedia - Social and Behavioral Sciences*, Vol. 182, 2015, pp. 163-172. https://doi.org/10.1016/j.sbspro.2015.04.752

[12] R.S. Anosov, S.S. Anosov, and I.Yu. Shakhalov, "Conceptual model of information technology security risk analysis", *Cyber Security Issues*, Vol. 2, No. 36, 2020, pp. 2-10. https://doi.org/10.21681/2311-3456-2020-2-02-10