

# Development of Models and Methods of Data Analysis for Enhancing Efficiency of the Processes of Quality Management Systems

Alla Vladimirovna Kuzminova<sup>1</sup> & Valeriy Valentinovich Gurov<sup>1</sup>

<sup>1</sup> National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409, Kashirskoe highway, 31, Moscow, Russia

Correspondence: Alla Vladimirovna Kuzminova, National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409, Russia, Moscow, Kashirskoe highway, 31, Russia E-mail: AVKuzminova@mephi.ru/VVGurov@mephi.ru

Received: January 9, 2015

Accepted: March 20, 2015

Online Published: June 25, 2015

doi:10.5539/mas.v9n8p429

URL: <http://dx.doi.org/10.5539/mas.v9n8p429>

## Abstract

This article deals with the education system, along with the other systems, which should include monitoring, measurement and analysis of the processes required to improve the quality management system. In this system, a student is considered as a product being created together with its input and output parameters and characteristics. Due to investigation of the nature of relations between the training process parameters, there are being developed methods and mathematical models that describe regularities of the system in order to enhance efficiency of quality management processes in education.

**Keywords:** Academic Success, Cluster Analysis, Regression Analysis, Maximum Likelihood Estimation, Principal Components Method, Levels of Quality

## 1. Introduction

Avalanche-like increasing flow of information makes it difficult for a contemporary person to qualitatively accept the necessary data, to process, comprehend, preserve, and create new knowledge.

In a current situation of information crisis, in order for a person to find his/her place in the sphere of material and socio-cultural production, special load must be transferred to the sphere of continuous training and education.

Teaching a person to navigate in this rapidly changing world requires time, but a significant part of the knowledge obtained, as well as the skills in its processing a person receives in the course of study in an institution of higher education. In this regard, the task of selecting applicants able to master more and more considerable volumes of scientific and technical knowledge and successfully complete training programs is faced.

In accordance with the ISO 9000 global standards (ISO Standards; State Standard of the Russian Federation, 2006), an organization shall carry out monitoring, measurement and analysis of the processes necessary to improve the system of quality management. Increase in recording and analysis of the quality of human resource development is possible if based on the established requirements for the evaluation of educational processes described in the National Standard GOST R 52614.2-2006, State Standard of the Russian Federation, 2006 (State Standard of the Russian Federation, GOST R 52614.2-2006). This standard includes guidelines for institutions providing educational services; it proclaims quality management systems in education. An information level of quality analysis system (Recommendations on Standardization, 2002, <http://www.complexdoc.ru/ntdtext/541946/>; The ISO/IEC/IEEE 42010 Website, <http://www.iso-architecture.org/ieee-1471/>; Cheremnikh, Semenov, & Ruchkin, 2003) is being developed. Parameters of product's conformity with the levels of quality required may serve as the input data for such analysis. While considering a university student within the educational system as a product being created, the education quality data in the form of examination marks for each term, and the psycholinguistic parameters of their written language, are analysed (Kuzminova, 2013; Sticht, 1973).

Problems of education quality are solved while measuring, analysing and improving processes of education system. These educational institution processes may be demonstrated through students training processes, determination of the process quality, and final evaluation of the achievement rate with awarding academic degree

to a graduating student in accordance with his/her diploma.

To enhance the efficiency of educational process, on the agenda there will be certainly raised a question concerning an individual tutor-navigation for a trainee, which depends not only on the features of his/her specialization and scores obtained during training, but also on the psychological personality characteristics, such as responsibility, purposefulness, ability to work and create a new theoretical or practical product.

To address one of the priority problems concerning formation of mechanisms for education quality evaluation, systems of evaluating vocational training quality are to be developed. Urgency of the problem under consideration is confirmed by a significant amount of studies performed in this area (Marukhina, & Berestneva, 2003; Educational Testing Service ETS, <http://www.ets.org/gre>). It is also determined by the necessity of excellence in a reduction in the number of applicants in connection with the demographic problems in the country. These processes are taking place against the backdrop of the increasing amount of information that you want to learn. The solution to this problem is particularly important in the transition to a multi-stage system of higher education in Russia.

Thus, the development of models, methods and means of data analysis is quite urgent in order to enhance the efficiency of quality management system processes, as well as the study of predicting methods for specialist training quality.

The analysis carried out has shown availability of a large variety of research efforts that are similar in theme. However, predicting of academic success (AcSuc) itself did not receive sufficient consideration. The reason for this is that the research issue considered in this work belongs to the interdisciplinary problems, and has not been so far clearly specified.

In this article, we will focus on the following issues:

- identification of the most significant factors affecting the determination of academic success levels;
- construction of mathematical models that makes it possible to identify availability of different-class data domains by means of investigated parameters of the educational process quality indicators;
- study of models of predicting academic success levels as an indicator of the educational process quality;
- development of predicting methods for the levels of mastering training programs by trainees;
- experimental test of the developed models and methods.

To address these issues, an analysis of the expert information data (Giarratino, & Riley, 2007; Muromtsev, 2005; Jackson, 2001) of the study sample was carried out, and on their basis academic success indicators were invented (Bogomolov, et al., 2009). Simultaneously, texts of the trainees' written works were considered. The analysis of syntactic parameters of the written language (Luria, 2002; Popov, Y., et al., 2001; Baranov, 2003; Flesch, 1948) was performed, which resulted in development of information text parameters. Analysis of the developed mathematical predicting models of academic success levels as an indicator of specialists training quality was carried out. Using computer-based methods of information processing, imaging of the results obtained during the study of constructed mathematical models of the boundary between the predictive levels was carried out. When creating an information system for determination of specialists' quality levels, these data are considered to be the system's input data. In order to determine the prognostic level of quality, mathematical models of success and information text model are created.

## 2. Methods

The purpose of this article is to describe the development of means that would contribute to the enhancement of quality management system efficiency in education through the predicting of the level of specialists training quality. This work demonstrates mathematical model for determining threshold between the predictive levels of specialists training quality and methods developed for their application. Models were created using methods of regression analysis and maximum likelihood estimation.

### 2.1 Construction of the Model for Classification of Academic Success Levels

The study of a complex system, as a rule, requires its preliminary partitioning into subsystems and determination of boundaries between them. There are many methods for solving this problem, ranging from classical statistical methods to methods of nonparametric statistics and neural networks. Each method has its advantages and disadvantages. To solve this problem, we propose a new approach that allows for the set output parameters of the process to obtain optimal in terms of established criteria boundary separating objects within the system.

The general scheme of solving the problem of the classification of quality levels with the use of this approach is

described as follows. Let us assume that we have a system, each element of which is characterized by a set of parameters  $\{W, X\}$ . Within the first subset of parameters  $\{W\}$ , using the Spearman rank correlation, a selection of the weightiest parameter  $W_p$  is carried out. Further, division of the system under investigation into groups is performed according to the dedicated parameter, using cluster analysis and applying the nearest-neighbour method. This creates clusters  $\{C_m\}$ , relating to different quality categories  $\{QC_m\}$ , which are considered in terms of the quality management system.

For the second subset of input parameters  $\{X\}$  of a certain process of the system in question, carried out is the typologization of its elements on the basis of their belonging to a particular cluster  $\{C_m\}$  found, defined as a class of the system quality.

As a result, predictive models of separating parameters  $X_k$  with a glance to  $W_p$  are being constructed – models of boundaries between the selections of points. Experimental test of the predictive models developed is conducted. We analyse the results, and choose the best according to certain parameters mathematical model of the boundary between predictive levels of academic success – functionality  $Q$ .

For this problem, as the input parameters of the educational process, exam results for all examinations passed during the educational period at university and parameters of penscripts obtained after the written entrance exams to university in the Russian language are selected.

The next step is the development of principles for analysing indicators of university educational process. For this purpose, a model of academic success classification is developed, as well as the models of predicting levels of academic success. When evaluating quantitative indicators of determination of the trainee's academic success levels, one should not be oriented on the trainee's grade point average obtained during the entire education period. It cannot serve as an objective indicator for evaluating the success of mastering the material during the training period, since it can be affected by the personal attitude of a teacher-expert to a particular student or by the student's individual attitude manifested to the subject ("clear – not clear") or emotional attitude to the teacher ("like – do not like"). Therefore, assignment of a trainee to the number of highly successful or slightly successful students on the basis of this indicator is impractical.

As a replacement for this seemingly obvious indicator, we propose to use 5 indicators of academic success and 3 formulas of academic success for the entire training cycle during  $n$  terms. For example, the indicator of academic success, describing GPA of the  $i$ -th trainee for the entire training cycle, is represented by the following formula:

$$A_i = \frac{\sum_{j=1}^n S_j}{\sum_{j=1}^n K_j}, \text{ in which the amount of scores } S_j, \text{ obtained during the } j\text{-th term } (j=1, \dots, n) \text{ for } K_j$$

exams is calculated. After analysing indicators of academic success with the help of Spearman rank correlation, the best of these indicators is selected.

For the further analysis of the selected indicators of academic success, the cluster analysis of time series of scores obtained by graduates as a vector of  $n$ -dimensional space points:  $\vec{W}_j = (W_1, W_2, \dots, W_n)$  is used. For this, sorting of expert evaluations is conducted through the method of cluster analysis, for which the parameters of cluster analysis (Kim, J., et al., 1989): Euclidean distance and nearest-neighbour method (single binding) are selected. As a result of such classification, clusters characterising different classes of academic success are

formed. To determine the parameters of the statistical and cluster analysis, the Statgraphics Centurion Program (Statgraphics: The Statistical Program) was used. As a result of partitioning the entire set of points  $W$ , clusters with different, in general case, number of elements are formed. These elements belong to different areas of quality categories – levels of academic success  $\{QC_m\}$ .

Thus, the construction of the model for classification of AcSuc levels with the implementation of a trajectory model (cluster-analysis) of academic success is carried out. By its application, it is possible to identify a dynamic component in the determination of academic success.

## 2.2. Mathematical models of Different-Class Data Separation

As a result of the analysis of sample distribution of random variables – developed transformational parameters of texts  $(x, y, z)$  – it becomes possible to use them in solving the task of searching for the classifying boundary of AcSuc.

Search for the boundary of different-class data separation can be performed using a variety of well-known classification models. In our work, we consider construction of a classifier using the method of multiple linear regression MR and the maximum likelihood method (MLM) – probabilistic model PM.

In the first classification model of MR, as a regression function applied is the function of three variables  $f(x, y, z, \vec{A})$  as a polynomial function of degree  $s$  ( $s \in N$ ), where  $\vec{A} = \{A_{s00}, \dots, A_{001}, A_{000}\}$  – are the required model parameters, which are set by the least-squares method. In the description of the model using three predictors, multiple polynomial regression model  $Qrg$  of the degree  $s$  is applied, which looks as follows (1):

$$Qrgs = f(x, y, z, \vec{A}) = A_{000} + A_{100}x + A_{010}y + A_{001}z + A_{200}x^2 + A_{110}xy + A_{101}xz + A_{020}y^2 + A_{011}yz + A_{002}z^2 + A_{300}x^3 + \dots + A_{421}x^4y^2z + \dots + A_{00s}z^s, \quad (1)$$

If you take recognition error as a criterion, i.e. the ratio of correct answers to their total number, which determines the quality of recognition, preference should be given to a suitable embodiment of the  $j$ -th regression model of  $MR_j$ . The study sample was obtained randomly from the entire assembly of data on the university trainees.

In the event of applying the second classification model, a probabilistic model PM, the educational process data  $(x, y, z)$  obtained are considered as random variables. The performed analysis showed that the resulting sample was normally allocated by the developed psycholinguistic text parameters. For the problem of estimating entire assembly parameters by sample data, there was used one of the methods of its solving demonstrated in the work (Kendall & Stewart, 1973). Mathematical model for determining the prognostic quality level is created by means of the maximum likelihood method application. Density of the normal probability distribution of a continuous

random variable for one-dimensional case is described as follows:  $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ , where  $a$  [alfa] – is the mathematical expectation of a continuous random variable,  $\sigma$  [sigma] – is the root-mean-square deviation of normal distribution.

And given that the function of input data is a function of three arguments, density of its distribution is constructed in the following form:

$$P(x, y, z) = \frac{1}{(\sqrt{2\pi})^3 \cdot \sigma_1 \cdot \sigma_2 \cdot \sigma_3} \cdot \exp\left\{-\frac{(f_1 - \theta)^2}{2\sigma_1^2}\right\} \cdot \exp\left\{-\frac{(f_2 - \theta)^2}{2\sigma_2^2}\right\} \cdot \exp\left\{-\frac{(f_3 - \theta)^2}{2\sigma_3^2}\right\}$$

where  $f_j = [x - \bar{x}] \cdot \cos(\alpha_j) + [y - \bar{y}] \cdot \cos(\beta_j) + [z - \bar{z}] \cdot \cos(\gamma_j); (j = 1, 2, 3)$ ,  $\cos^2(\alpha_j) + \cos^2(\beta_j) + \cos^2(\gamma_j) = 1;$

$$\cos(\alpha_k) \cdot \cos(\alpha_j) + \cos(\beta_k) \cdot \cos(\beta_j) + \cos(\gamma_k) \cdot \cos(\gamma_j) = 0; (k \neq j)$$

Unknown parameters  $\cos(\alpha_j), \cos(\beta_j), \cos(\gamma_j); (j = 1, 2, 3)$  shall be determined by the principal components method.

Evaluation of the parameter  $\theta$  [teta] was carried out.

Having indicated:  $\sum_n \{[x_n - \bar{x}] \cdot \cos(\alpha_j) + [y_n - \bar{y}] \cdot \cos(\beta_j) + [z_n - \bar{z}] \cdot \cos(\gamma_j)\}^2 = S_j^2 \cdot N$  and using method of

maximum likelihood, there has been obtained an optimization problem:  $\sum_j \left( \frac{s_j^2 \cdot N}{\sigma_j^2} + 2 \cdot \ln(\sigma_j) \cdot N \right) \rightarrow \min$ , which

solution is found as follows  $\sigma_j = S_j$

Given that the observation vector consists of independent random variables, obtained is a partition of the likelihood function into contributions of individual observations. For normal samples of two classes of transformational text parameters  $(x, y, z)$ , in view of belonging to different levels of quality:  $P_1(x, y, z)$  and  $P_2(x, y, z)$ , densities of probability distributions are determined.

For the first level it is obtained as follows:

$$P_1(x, y, z) = \frac{1}{(\sqrt{2\pi})^3 \cdot \sigma_1 \cdot \sigma_2 \cdot \sigma_3} \cdot \exp\left\{-\frac{\{[x - \bar{x}] \cdot \cos\alpha_1 + [y - \bar{y}] \cdot \cos\beta_1 + [z - \bar{z}] \cdot \cos\gamma_1\}^2}{2 \cdot \sigma_1^2}\right\} \cdot \exp\left\{-\frac{\{[x - \bar{x}] \cdot \cos\alpha_2 + [y - \bar{y}] \cdot \cos\beta_2 + [z - \bar{z}] \cdot \cos\gamma_2\}^2}{2 \cdot \sigma_2^2}\right\} \cdot \exp\left\{-\frac{\{[x - \bar{x}] \cdot \cos\alpha_3 + [y - \bar{y}] \cdot \cos\beta_3 + [z - \bar{z}] \cdot \cos\gamma_3\}^2}{2 \cdot \sigma_3^2}\right\}; \tag{2}$$

Similarly, we obtain expression for the second level of quality  $P_2(x, y, z)$ .

Evidently, in the area of class separating hyperplane propagation, we can observe equality of probabilities of belonging to them. As a result, the surface – boundary separating the ellipsoids of data diffusion, which belong to different levels of quality is determined.

Based on the statistical analysis data of three-dimensional ellipsoids of diffusion and the principal components method, we carry out identification of the boundary separating two ellipsoids of data diffusion from the condition of probabilities equality at the boundary. A polynomial of the following form is obtained (3):

$$QQ = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{13}xz + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} \tag{3}$$

At  $QQ = 0$  polynomial (3) is represented as an equation of a second order surface in space  $(x, y, z)$ . Unreduced form of the second order surface equation is reduced to a canonical form by classical manner. Based on the

analysis of the canonical equation type, it is possible to determine the type of surface that we will call a functional of the academic success.

### 2.3 Method for studying processes of education quality management and its approbation

Based on the developed models of predicting levels of training quality, a methodology for conducting studies of process parameters that require division of system elements into classes is developed.

Approbation of the developed models, methods and data analysis was carried out on the example of data on the NRNU MEPhI students. For this purpose, a representative subsample of students belonging to the categories of "graduates" and "students expelled for academic failures" was taken.

Analysis of indicators of academic success confirmed that the best approximation to the median estimates among the indicators of academic success was a highlighted parameter, a truncated middle-terminal score  $B_j$ , which is determined for the entire cycle of student learning, after deduction of one the best and one the worst mark in each term  $J$ .

In accordance with the developed algorithms for constructing mathematical models, a cluster analysis of the selected parameter  $B_j$  was performed. The "nearest neighbour" method was implemented. An experimental

study of the model of classifying levels of academic success for the highlighted parameter  $B_j$  was carried out. As a result of the cluster analysis of university graduates sample, out of 2446 points of baseline data on examination scores identified are 2 large clusters belonging to different levels of AcSuc: categories of students are defined as "highly successful" (32%) and "slightly successful" (55%). In addition, six small clusters containing 1-2 elements are specified. These 13% of students make up the category of "moderately successful" students.

Analysis of the developed mathematical models – regression models of MR (1) and with implementation of MLM – PM models (3) is conducted.

### 2.4 Regression Models of Determining the Level of Quality

As the analysis of regression models to the 4th degree inclusive has shown, conducted stepwise method of improving regression models does not give any significant improvement. It is natural to assume that for the higher degrees of polynomial it is possible to obtain a model, which will more accurately explain the share of the dependent variable variation. However, this increases the bulkiness of the model itself, increases the complexity of the regression model, which is determined by the number of predictors included in it, and increases the complexity of the calculations.

For the experimental sample obtained are MR – classification regression models  $Qrgs$  for different degrees  $s$  of the polynomial. Thus, for the degree  $s=1$  the following model is defined:  $Qrg1 = 1.1 - 1.6x + 1.3y - 0.03z$ .

And for the fourth-degree polynomial ( $s=4$ ) a regression model of the following form is developed:

$$\begin{aligned} Qrg\ 4 = & -7390.1 + 20291.1x + 951.7y - 1899.0z - 20268.6xx - 2071.6xy + 2949.9xz + 602.3yy - \\ & - 389.7yz + 369.1zz + 9314.8xxx - 1163.6yyy - 43.7zzz + 1687.3xyz + 2010.4xyy - 292.7xzz - \\ & - 70.1yzz - 633.5xxy - 1951.5xxz - 930.4yyz - 1694.3xxxx + 231.9yyyy + 0.5zzzz + 520.1xxxz + \\ & + 704.5xxxy + 76.1yyyx + 174.6yyyz + 32.0zzzx - 12.2zzzy - 764.4xxyz + \\ & + 272.7yyxz + 61.6zzxy - 838.3xxyy + 24.0xxzz + 21.3yyzz \end{aligned}$$

For regression models, their stepwise changing is carried out by the method of successive elimination of variables from the model, which criterion of p-value is the largest, and the p-value  $\geq 0.05$ .

Finally, completed at a certain  $k$  step ( $stp_k$ ) regression models of the classification boundaries are obtained. Thus,

for the linear polynomial regression of the 4<sup>th</sup> degree, a calculated by 31 steps model looks as follows:

$$Qrg_4(stp\ 31) = 1.2 - 1.5xy - 0.1xzz + 1.1yyyz$$

As a result, this mathematical model allows describing the relationship between the function Qrg and 34 independent predictors.

For the selected regression model  $Qrg^3$ , based on the analysis of about 100 models, the pairs of error functions of the I  $O_1$  and II  $O_2$  types are experimentally determined.

Based on their analysis, it is found out that the developed model  $Q_{19}$  – is the model with the highest sensitivity (100%), which gives true result at presence of positive outcome (it reveals positive examples better than other models):

$$Q_{19} = -6.8x^2 + 3.6y^2 - 0.2z^2 + 5.6xy + 1.2xz + 2.5yz + 14.4x - 18.3y - 3.1z - 0.8$$

And the  $Q_{22}$  model – is the model with the highest specificity, which gives true result at the presence of negative outcome. This model identifies negative examples rather well:

$$Q_{22} = -12.4x^2 + 6.6y^2 - 0.5z^2 + 10.3xy + 2.3xz + 4.6yz + 26.3x - 33.5y - 5.8z - 2.3$$

### 2.5 Probabilistic model for determining the level of quality

Densities of text indicators distribution  $(x, y, z)$  for two samples are defined in view of belonging to different levels of quality. For this purpose, the statistical analysis and the method of principal component analysis are carried out. Identification of boundaries out of the condition of probability equality at the boundary separating two ellipsoid of data diffusion ("slightly successful" students and "unsuccessful" trainees) is conducted. Obtained is a polynomial of the form (3), and the unreduced form of the second order equation of surface is reduced to the canonical form.

As a result, equation (4) is developed, which determines a one-sheet hyperboloid of revolution shown in Figure 1.

$$QQ_{2\setminus 3} = 1.43 + H_1^2 - H_2^2 - H_3^2, \quad (4)$$

where  $H_1, H_2, H_3$  – are the linear combinations of text parameters  $(x, y, z)$   $H_i = k_i x + l_i y + m_i z + n_i$ , a

$k_i, l_i, m_i$  ( $i=1, 2, 3$ ) – coefficients obtained as a result of the second order surface reduction to a canonical form.

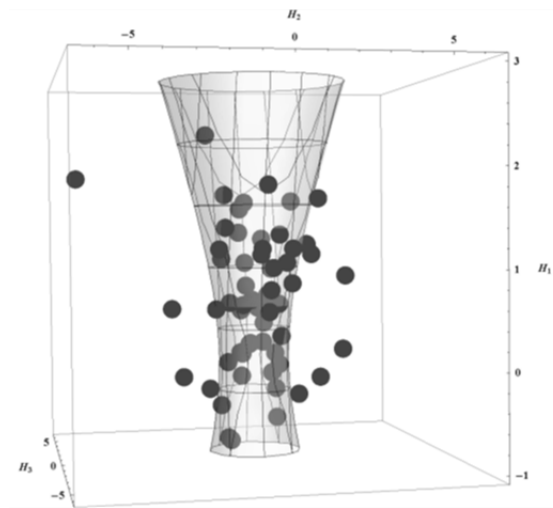


Figure 1. Classification boundary "Unsuccessful – Slightly Successful"

In such a way, a mathematical PM model of academic success at university training for "unsuccessful" and "slightly successful" graduates was constructed. At the boundary of class division, equality of probabilities of belonging to them is observed. For definiteness, it is assumed that the points at  $QQ > 0$  are located inside the domain, and at  $QQ < 0$  – they are located outside the domain.

Calculations and visualization of the study results were carried out with the help of the symbolic computation software system WolframMathematica (Website of the Wolfram Company. URL: <http://wolfram.com/resources/>).

### 3. Analysis of Mathematical Models

Due to the analysis of the developed mathematical models of predicting academic success levels, a PM model, which was more effective in predicting compared with a MR model, was selected.

To check the adequacy of the developed models, methods and techniques, a set of the best essays of potential applicants written by competitions winners in Russian literature and school medal winners was subjected to a screening analysis.

Check of the hypothesis about the type of distributions for transformational indicators of texts  $(x, y, z)$  of the

"Russia Medal Winners" sample (fitting criterion  $\chi^2$  and Kolmogorov-Smirnov criterion) has shown that we cannot reject the hypothesis of normal distribution at a significance level of 10.

Further, assuming that the works could be written by medal winners-applicants at entrance exams to the NRNU MEPhI, with respect to a random sample of medal winners, an approbation of the developed models and techniques was carried out.

The probability of applicants belonging to one of the classes of academic success was determined: to the level of successful students or to the level of unsuccessful students (slightly successful and unsuccessful). Thereto, a PM model developed for these levels and function of academic success  $QQ_{2/3}$  were used.

It is determined that some of these AcSuc functions of the Russia Medal Winners belong to the predictive class of "unsuccessful" students at training in the NRNU MEPhI. And only 75% of the Russia Medal Winners



obtained values of the function  $QQ > 0$ , which allowed predicting for them a successful completion of training in the NRNU MEPhI. The performed analysis of practical data showed that those were 83% of the applicants-medal winners. That is, we see that the predicting accuracy by this model is about 90%.

#### 4. Results

Based on the implementation of principles of the national standard for quality management system in education, methods and data analysis tools for enhancing efficiency of the educational process are developed:

1. Based on consideration of the basic concepts of national education development and works of foreign researchers, relevance of construction and analysis of efficient ways expected to improve the quality of higher education and their relation to the problem of excellence is shown.
2. The study of the "trainee" subsystem's output parameters is carried out, and an original method of beam projection classification is developed, confirming existence of systemic connection between the selected output parameters and providing for determination of classification boundaries availability between the subsystem objects.
3. The quantitative indicators of academic success are developed. It is determined that the dynamic analysis of academic success shall be carried out based on the analysis of the truncated middle-terminal scores.
4. Original models of academic success classification are created. In particular, the models of predicting levels of academic success through regression and probabilistic (statistical) approaches are developed.
5. Technique for separating surface acquisition is developed. For the first time, the functional for probabilistic predicting of academic success levels in the form of a second-order equation of the surface – the classification boundary between different-level data is acquired.
6. The verification of developed models, methods and techniques for probabilistic predicting of academic success levels in a sample of the best essays written by competition winners in Russian literature and medal winners is conducted. The adequacy of the chosen model has been proven by example of "Russia Medal Winners" and a set of applicants-medal winners, the NRNU MEPhI students. The result predicted by the model coincided with the actual result by 87%.
7. Classification boundaries between the levels of training quality are constructed. The constructed models of predicting levels of academic success ("highly successful – slightly successful – unsuccessful") has shown that the results predicted by the developed mathematical models of PM for different levels of academic success coincided with the actual results by 70 to 90% of cases.

#### 5. Discussion

At carrying out of the system study of the "trainee" subsystem's output data, there have been developed methods and data analysis tools for enhancing the efficiency of the quality management system processes in education.

The most significant factors affecting the determination of levels of academic success are singled out; a model that allows to determine the presence of different-class data domains is constructed; mathematical models of predicting levels of academic success as an indicator of the educational process quality are developed and investigated; a method of predicting the level of training program mastering by students is developed; the experimental test of the models and techniques developed is executed. Designed functional of determining the level of training quality, boundaries separating different-type categories of students will provide for predicting quality of the educational process of the main stages of multi-level education (Bachelor, Specialist, Master), which will be expressed in cost, time, and health savings.

For evaluation of examination marks provided by expert teachers, studied is their changing in the process of acquiring knowledge for the entire training period. Based on consideration of correlation between the individual results for all examinations, both oral and written, its existence is explained by availability of a psychological component.

Regression models that to the best advantage identify "high levels of quality" and "low levels of quality", but have a low interpretation degree of variables in the models are obtained. Aggregate analysis of the obtained regression models (Gurov, & Kuzminova, 2012) is carried out. A statistically significant relationship between the variables at a confidence level of 95.0% is identified. Statistic R-Squared value demonstrated that the models obtained cannot explain more than 25% of the variability in the MR models.

In contrast to the regression models, the mathematical model developed by application of maximum likelihood method of PM allows predicting these levels to 85% of cases.

In the course of further work there shall be given the answers to the following questions:

- What other separating surfaces can be formed?
- What is the specificity of data belonging to different levels besides their affiliation to a particular cluster of academic success?

## 6. Conclusion

Thus, we developed mathematical models of data analysis using different methods; evidenced the great practical value of created functional; constructed mathematical models of classification boundaries between the predictive levels of academic success through the application of maximum likelihood method. Implementation of the developed model of PM allows classifying levels of quality, predicting levels of training quality at a higher level compared to the other models considered, enhancing efficiency of quality management system processes. To increase the quality of the developed models, it is permissible to explore the additional parameters of texts or bring into consideration other parameters that affect the levels of training quality.

## Acknowledgements

The authors highly appreciate creative abilities of the Professor, Doctor of Technical Sciences Popov, Y.A. from the "Computer Systems and Technologies" Department of the National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), owing to whom this research work was carried out, and pay tribute to his blessed memory. Authors value highly professional support by the Candidate of Technical Sciences Assistant Professor Zhigirev, N.N., the chief leading researcher of the Science Research Institute of Economics and Management in Gas Industry in study of multidimensional data, a specialist in the field of applied mathematics.

## References

- International Organization for Standardization. (n.d.). *ISO Standards*. Retrieved January 8, 2014, from [http://www.iso.org/iso/ru/iso\\_9000](http://www.iso.org/iso/ru/iso_9000)
- Recommendations on Standardization of the Russian Federation. R 50.1.028-2001 Continuous Acquisition and Life Cycle Support Technology. Methodology of functional modelling.* (2002). Moscow. Retrieved January 8, 2014, from <http://www.complexdoc.ru/ntdtext/541946>
- Cheremnykh, S., Semenov, I., & Ruchkin, V. (2003). *Structural analysis of systems: IDEF-technology*. Moscow: Finance and Statistics.
- Kuzminova, A. (2013). Application of text analysis methods at levels of quality forecasting. *Quality. Innovation. Education*, 12, 27-29.
- Sticht, T. (1973). Research towards the design, development and evaluation of a job-functional literacy training program for the US Army. *Literacy Discussion*, 4, 339-369.
- Marukhina, O., & Berestneva, O. (2003). Forecasting of the success of students training based on non-uniform sequential pattern recognition procedure. *Computer modelling: Proceedings of the IV International Scientific and Engineering Conference in St. Petersburg*. Retrieved January 8, 2014, from <http://berestneva.am.tpu.ru/Papers/KONF2009/PITER-2003.doc>.
- Giarratino, J., & Riley, G. (2007). *Expert Systems: Principles of Design and Programming*. Moscow: "Williams Publishing Company", LLC.
- Muromtsev, D. (2005). *Introduction to the Technology of Expert Systems*. St. Petersburg: SPb ITMO University.
- Jackson, P. (2001). *Introduction to Expert Systems*. Moscow: "Williams" Publishing Company.
- Bogomolov, A., et al. (2009). Forecasting of Trainees' Academic Achievements in Special Disciplines on the Basis of Regression Equations. *Herald of the Volga region, Humanities Series, Proceedings of the Higher Educational Institutions*, 1, 124-132.
- Luria, A. (2002). *Writing and Speaking. Neurolinguistic Researches*. Moscow: The Academy.
- Popov, Y., et al. (2001). *Psycholinguistic technologies of personality assessment*. Moscow: PAIMS.
- Baranov, A. (2003). *Introduction into Applied Linguistics*. Moscow: Editorial URSS.
- Flesch, R. (1948). A New Readability Yardstick. *Applied Psychology*, 3, 221-233.

- Kim, J., et al. (1989). *Factor, Discriminant and Cluster Analysis*. Moscow: Finance and Statistics.
- Statgraphics: The Statistical Program*. (n.d.). Retrieved January 8, 2014, from <http://statgraphics.com/downloads.htm>.
- Kendall, M., & Stuart, A. (1973). *The advanced theory of statistics*. Moscow: Science.
- Gurov, V., & Kuzminova, A. (2012). Analysis of regression models of system processes indicators, which study requires division of its elements into classes. *Natural and Technical Sciences*, 2, 314-318.
- Educational Testing Service ETS*. (n.d.). Retrieved January 8, 2014, from <http://www.ets.org/gre>.
- State Standard of the Russian Federation. GOST R 52614.2-2006 Quality Management Systems. Guidelines on Application of GOST R ISO 9001-2001 in Education. (2006). *Federal Agency on Technical Regulating and Metrology Website*. Retrieved January 8, 2014, from <http://www.gost.ru/wps/portal/pages.CatalogOfStandarts>.
- The ISO/IEC/IEEE 42010. (n.d.). Retrieved January 8, 2014, from <http://www.iso-architecture.org/ieee-1471>.
- Website of the Wolfram Company*. (n.d.). Retrieved January 8, 2014, from <http://wolfram.com/resources>.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).