Innovative Methods of Assessing the Academic Success of Students in Higher Education Institutions

Rostyslav Yurynets [1], Zoryna Yurynets [2], Nataliia Danylevych [2]

[1] *Lviv Polytechnic National University, Lviv, Ukraine*
[2] *Ivan Franko Lviv National University, Lviv, Ukraine*

Abstract

In the article use innovative possibilities of assessing and predicting the academic success of students in higher education institutions are investigated. The proposed multinomial logistic regression model allows to assess and predict the success of students based on the obtained estimates and taking into account various factors. The multinomial logistic regression model is one of the best tools for determining the quality of training for students, proceeding from the quantitative (intellectual) and qualitative (behavioral) factors. The following factors were used in the study: an assessment of the student's presence in the classroom; student's grade in mathematics; student's grade in computer science; assessment of the student's living conditions. The use of regression analysis methods allows identifying implicit relationships between elements of the educational process and solve many problems related to the analysis and prediction of academic success of students. A multinomial logistic regression model confirms that assessment of the student's presence in the classroom; student's grade in mathematics; student's grade in computer science; assessment of the student's living conditions are important indicators that reflect the academic success of students in higher education institutions. It is possible to predict individual learning outcomes, evaluate student actions, and assess factors related to success, create an individual learning plan through the constructed model. This knowledge is significant for university administration and teachers who want to improve the quality of teaching and ensure better learning outcomes.

Keywords [1]

Academic Success, Students, Higher Education Institutions, Teachers, Multinomial Logistic Regression Model, Factors

1. Introduction

Educational data mining plays an important role in the development of the learning environment. Modern higher education institutions operate in a highly competitive and complex environment. The main problems faced by higher education institutions are: analysis of learning outcomes, quality education, evaluation of student performance, strategic decision-making, and formation of human capital at all. Universities need to implement student intervention plans to address the challenges that students face during their studies.

Different methods of machine learning are used to understand and overcome the main problems. Various innovative methods and methods of machine learning are used to understand and overcome the main issues. Modern development of society requires the use of new innovative methods of assessing students' educational levels in higher education institutions, evaluating of the knowledge gained inside the higher education, training of students in various subjects, in particular information systems in management. These methods will allow future professionals to be more competitive in the labor market. Innovative methods in higher education are characterized as technologies based on innovations: organizational (related to the optimization of educational conditions), methodological (updating the

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: rostyslav.v.yurynets@lpnu.ua (R. Yurynets); zoryna_yur@ukr.net (Z. Yurynets); danylevychnatali@gmail.com (N. Danylevych)
ORCID: 0000-0003-3231-8059 (R. Yurynets); 0000-0001-9027-2349 (Z. Yurynets); 0000-0001-9906-1492 (N. Danylevych)

content of education and improving its quality); technical (related to the use of new tools of artificial intelligence in the field of informatics).

Machine learning methods play an important role in predicting the success of students at risk and dropout, thereby improving learning outcomes. Machine learning and statistical methods for educational data are needed to identify significant patterns that improve students' knowledge and the learning process in higher education institutions. Predicting student achievement at the grassroots level and beyond helps universities develop intervention plans in which the executive heads and university professors are beneficiaries of student performance prediction plans. Predicting academic success of students provides excellent benefits for improving student retention rates, targeted marketing, and overall student performance, as well as for effective enrollment management, graduate management.

The hypothesis is that in order to achieve student's academic success and qualifications in management information systems, it is necessary that internal elements be considered as an integral part of intellectual and behavioral factors. The proposed multinomial logistic regression model allows university administration and teachers to assess and predict the success of students based on the obtained estimates and taking into account various factors. The following factors were used in the study: an assessment of the student's presence in the classroom; student's grade in mathematics; student's grade in computer science; assessment of the student's living conditions. Assessing the academic success of students in the discipline of information systems in management is the ultimate task.

The model also allows predicting the individual learning outcomes, to evaluate student actions and differences between actions, to assess factors related to success, to create an individual learning plan through the constructed model. The theoretical and methodological basis of the paper is the scientific and research works of domestic and foreign academicians, which reveal the issues of educational management, modeling student performance.

The survey rests on the following general and specific methods: a systematic approach to the study of machine learning methods to improve the quality of the educational process in higher education institutions; a deductive method to solve the problem of systematization of the factors that affects the academic success of students; multinomial logistic regression analysis as a machine learning method to solve the problem of forming the model of the influence of factors on the assessing the academic success of students and get an information about the level and quality of the learning in higher education institutions; analytical method for assessing the state of academic success of students in higher education institutions.

2. Literature review

Analysis of the academic achievement of students at all levels of training is a prerequisite for the effective functioning of the management system of the educational process. The process of managing human capital preparation involves a wide range of interdependent and complementary components. An important component is the analysis of the organization of the educational process and, foremost, the process of forming the academic success of students.

The relevance of methods of regression data analysis in educational systems is confirmed by a number of studies of learning processes.

Kovacic [1] analyzed the results of predicting student performance through machine learning methods. Demographics (education, employment, gender, status, disability, etc.) and course characteristics (course syllabus, course unit, etc.) were taken into account. The data were collected at the Open University of New Zealand. The main features influencing student success were identified through machine learning algorithms. Three main factors influencing student performance (course syllabus, course unit, and ethnicity) were identified. It is important to choose subjects that determine the success of students in the course program.

Lakkaraju [2] suggested identifying failed students or students who risk not graduating on time. To do this, he used the structure of machine learning. Using this structure, student data was collected from two schools in two districts. The study took into account five machine learning algorithms: Support Vector Machine (SVM), Random-Forest (RF), Logistic Regression, Adaboost and Decision Tree. These algorithms are evaluated using accuracy, response, and AUC for binary classification. Each student is evaluated based on a risk assessment evaluated according to the proposed classification. The author also paid important attention to the impact of student presence and involvement in the learning process. Teachers need to look for opportunities to strengthen presence and involvement of student with regard to learning process in higher education institutions. The results showed that Random Forest demonstrates the best performance.

Oyedeji and other scholars [3] have used machine learning techniques to analyze student performance as an element of the formation of human capital. The obtained results allow educational institutions to eliminate methods that cannot improve student performance. Their study took into account individual attributes, including the age of the student, their demographic distribution, individual attitude to the survey and marital status. The authors noted the need to include in the study of the academic success of students the student's living conditions. Various machine learning algorithms were used in the study. For a comparative analysis of productivity, three models have been identified: learning with a teacher using linear regression and in-depth learning.

Alhusban [4] used machine learning methods to measure and increase the dropout undergraduate students. They collected data from al-Bayt University students. The following factors were taken into account: gender, type of enrollment, enrollment marks, city of birth, marital status, nationality and subjects studied at stage k-12. The researchers used Hadoop and an open source platform based on machine learning. It has been found that the marks for entrance examinations significantly affect specialization. The researchers have been stated that students' social status has a significant impact on academic performance. This is of great importance for the development of higher education institutions and the maintaining human capital.

Our study took into account the ideas of scientists who emphasize the importance of the influence of factors (course programs, student's living conditions, student's presence in the classroom) on the academic success of students. The course program of the specialty "Management information systems" include various subjects. In our study, important subjects such as computer science and mathematics were selected. These subjects are the basis of the specialty "Management information systems".

3. Methodology

To assess the academic success of students, a relationship should be established between factors and academic success. We will take the student’s academic success indicator according to three values (good, satisfactory and unsatisfactory). Therefore, it is necessary to predict the success of students through a model of multinomial logistic regression. Logistic regression is used to predict the probability of an event by the value of a set of features.

Now, consider modeling assessments of students' academic success in order to predict their future condition. This means estimating a qualitative variable (intellectual factors) by several quantitative factors (behavioral factors). Discriminant analysis tools can be used to select the most informative quantitative variables. Logit regression allows determining the group of students' academic success. In addition, logistic regression makes it possible to consider the probability that a student will be classified as a certain group success.

The probability function is the basis of the method and expresses the probability density of the simultaneous appearance of the sample results $Y_1, Y_2, \dots, Y_n$ [5]:

$$L(Y_1,Y_2,\dots,Y_k;\Theta)=p(Y_1;\Theta)\cdot\dots\cdot p(Y_n;\Theta) \tag{1}$$

According to the method of maximum likelihood, the value of $\Theta=\Theta(Y_1,\dots,Y_n)$ [6], which maximizes the function L is taken to estimate the unknown parameter.

The calculation process is simplified by maximizing not the function L, but the natural logarithm $\ln(L)$. This is due to the fact that the maximum of both functions is achieved at the same values Θ [7]:

$$L^*(Y;\Theta)=\ln(L(Y;\Theta))\rightarrow\max \tag{2}$$

All multiple choice options are numbered in random order 0, 1, 2, ..., J [8]. The probability of occurrence of an option is described by a polynomial logit model

$$P(y_i = j) = \frac{e^{x_i \boldsymbol{b}_j}}{\sum_{j=0}^{J} e^{x_i \boldsymbol{b}_j}} \tag{3}$$

where b_j – unknown parameters that we will evaluate;

x_1 – assessment of the students’ presence in the classroom (behavioral factor);

x_2 – students' grade in mathematics (intellectual factor);
x_3 – students' grade in computer science (intellectual factor);
x_4 – assessment of the students' living conditions (behavioral factor).
The following notations are introduced here [9]:

$$\boldsymbol{b} = (b_0, b_1, \dots, b_m)^T,$$
$$\boldsymbol{x}_i = (1, x_{i1}, \dots, x_{im}), \tag{4}$$
$$\boldsymbol{x}_i\boldsymbol{b} = b_0 + b_1x_{i1} + b_2x_{i1} + \dots + b_mx_{im}$$

To identify the model (3), usually use the rationing $b_J = 0$ [10]. Then

$$P(y_i = j) = \frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{1+\sum_{j=0}^{J-1} e^{\boldsymbol{x}_i\boldsymbol{b}_j}}\ j = 0, 2, \dots, J\text{–}1, \tag{5}$$

$$P(y_i = J) = \frac{1}{1+\sum_{j=0}^{J-1} e^{\boldsymbol{x}_i\boldsymbol{b}_j}} \tag{6}$$

This is one of the features of building a polynomial logit model. According to this feature, only the coefficients of the first J dependencies $b_0, b_1, \dots, b_{J-1}$ are calculated, in accordance with (5) which the first J probabilities $P(y_i = 0)$, $P(y_i = 1)$, ... , $P(y_i = J - 1)$. The probability of choosing the last option $P(y_i = J)$ is not calculated, but determined separately using (6).

The coefficients of the model are estimated by numerically solving the plausibility equations. To write the plausibility equation itself (its logarithmic form) it is convenient to enter the variable d_{ij}, which becomes 1, if the *i-th* observation was chosen *j*-th alternative among (J + 1), and 0 – otherwise. Then for each i only one d_{ij} will be equal to 1.

Using the entered variable d_{ij} we write the logarithmic likelihood function [11, 12]

$$\ln L = \sum_{i=1}^{n}\sum_{j=0}^{J} d_{ij}\ln\left(\frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}\right) \tag{7}$$

Differentiating expression (7) by $\boldsymbol{b}_j$, we obtain a system of equations of maximum likelihood

$$\begin{aligned}\frac{\partial \ln L}{\partial \boldsymbol{b}_j} &= \sum_{i=1}^{n}\sum_{j=0}^{J} d_{ij}\frac{\partial}{\partial \boldsymbol{b}_j}\ln\left(\frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}\right) = \\ &= \sum_{i=1}^{n}\sum_{j=0}^{J} d_{ij}\frac{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}\left(\frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k} - e^{\boldsymbol{x}_i\boldsymbol{b}_j}e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\left(\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}\right)^2}\right)\boldsymbol{x}_i' = \\ &= \sum_{i=1}^{n} d_{ij}\left(1 - \frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}\right)\boldsymbol{x}_i' = 0,\ j = 0, 1, 2, \dots, J\text{–}1.\end{aligned} \tag{8}$$

The solution of this system, given that $\boldsymbol{b}_j = 0$, is carried out numerically using the Newton-Rafson method [13]. The computer implementation is arranged as follows: the values of the model that corresponds to the last of these alternatives become zero. In other words, if we want $\boldsymbol{b}_0 = 0$, instead of $\boldsymbol{b}_j$, then the data corresponding to the alternative number $j = 0$, must be entered last.

The Newton-Rafson method usually requires several iterations [14]:

$$\boldsymbol{b}_{t+1} = \boldsymbol{b}_t - \frac{\partial \ln L(\boldsymbol{b}_t)}{\partial \boldsymbol{b}}\left[\frac{\partial^2 \ln L(\boldsymbol{b}_t)}{\partial \boldsymbol{b}\,\partial \mathbf{b}'}\right]^{-1} \tag{9}$$

To implement the Newton-Rafson method requires a matrix of partial derivatives of the second order [14]:

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial \boldsymbol{b}_j \partial \boldsymbol{b}_i'} &= \frac{\partial}{\partial \boldsymbol{b}_i'}\sum_{i=1}^{n}\left(d_{ij} - \frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}\right)\boldsymbol{x}_i' = \\ &= -\sum_{i=0}^{n}\left(\frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k} - e^{\boldsymbol{x}_i\boldsymbol{b}_j}e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\left(\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}\right)^2}\right)\boldsymbol{x}_i'\boldsymbol{x}_i = \\ &= -\sum_{i=0}^{n}\left[\frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}\left(\boldsymbol{I}(j = 1) - \frac{e^{\boldsymbol{x}_i\boldsymbol{b}_j}}{\sum_{k=0}^{J} e^{\boldsymbol{x}_i\boldsymbol{b}_k}}\right)\right]\boldsymbol{x}_i'\boldsymbol{x}_i\end{aligned} \tag{10}$$

In the obtained expression, $\boldsymbol{I}(j = l)$ becomes 1 when $j = l$ and 0 in other cases.

4. Empirical results

The main data for the statistical analysis of the educational process are data on the success of undergraduate students at the economic faculty of Ivan Franko National University of Lviv (Table 1), where

x_1 – assessment of the students' presence in the classroom;
x_2 – students' grade in mathematics;
x_3 – students' grade in computer science;
x_4 – assessment of the students' living conditions.

Table 1
Data table for assessing academic success of students

№	Success in information systems in the management	Students' presence in the classroom	Grade in mathematics	Grade in computer science	Students' living conditions
7	1	8	7	7	2
8	0	8	6	6	2
9	2	10	7	8	3
10	0	8	7	6	1
11	1	7	8	7	2
12	1	7	8	6	3
13	0	8	6	7	3
14	2	10	8	8	2
15	0	7	6	6	3
16	1	7	7	6	3
17	2	10	7	7	3
18	1	8	8	7	2
19	1	8	6	8	2
20	0	8	5	8	2

Analysis of the data showed the presence of multicollinearity. Therefore, it is advisable to use the method of principal components. The main idea of the method is to replace highly correlated variables with a set of new variables between which there is no correlation. The new variables are linear combinations of the original variables

$$\boldsymbol{z} = \boldsymbol{x} \cdot \boldsymbol{a} \tag{11}$$

First, we need to normalize all the explanatory variables:

$$x_{ij}^{*} = \frac{x_{ij} - \overline{x_j}}{\sigma_{x_j}}, \quad i = 1,2,\dots,n; \; j = 1,2,\dots,m \tag{12}$$

Next, we need to calculate the correlation matrix

$$r = \frac{1}{n}(X^{*T}X^{*}) \tag{13}$$

Then we need to find the characteristic numbers of the matrix r from the equation

$$|r - \lambda E| = 0 \tag{14}$$

at the same time

$$\sum_{k=1}^{m} \lambda_k = m \tag{15}$$

where E is a unit matrix of size $m \times m$.

The eigenvalues λ_k $(k = 1,2, \dots, m)$ are ordered by the absolute level of the contribution of each main component to the total variance.

It is necessary to calculate the eigenvectors a_k by solving the system of equations

$$(r - \lambda E)a = 0 \tag{16}$$

Finding the main components (vectors) is determined by the formula:

$$Z_k = X^* \cdot a_k \tag{17}$$

As a result of calculations, characteristic numbers and eigenvectors have been found:

$$a = \begin{pmatrix} 0{,}77 & -0{,}01 & -0{,}18 & -0{,}62 \\ -0{,}43 & 0{,}71 & 0{,}05 & -0{,}56 \\ -0{,}42 & -0{,}71 & 0{,}16 & -0{,}55 \\ 0{,}24 & 0{,}07 & 0{,}97 & 0{,}01 \end{pmatrix} \tag{18}$$

$$\lambda = (0{,}53 \quad 0{,}72 \quad 1{,}03 \quad 1{,}72)$$

We find the connection of the dependent variable Y with the main principal components z_3 and z_4, defining all principal components and rejecting those that correspond to small values of characteristic roots. To do this, we build a model of multinomial logistic regression. The parameters of the obtained multinomial logit model in the STATISTICA environment are as follows:

Table 2

Parameters of the multinomial logit model

y - Parameter estimates (Spreadsheet1_1) Distribution : MULTINOMIAL Link function: LOGIT							
	Level of - Effect	Level of - Response	Column	Estimate	Standard - Error	Wald - Stat.	p
Intercept 1		0	1	-1,0001	0,074498	1,671322	0,18942
z3		0	2	-12,9008	9,00237	2,053606	0,151846
z4		0	3	25,0441	12,54538	2,964557	0,085108
Intercept 2		1	4	3,7657	2,16029	1,918160	0,173349
z3		1	5	-11,2274	6,94633	1,874960	0,179488
z4		1	6	19,7129	11,34079	1,809539	0,169254
Scale				1,0000	0,00000		

It can be concluded that the obtained estimates of coefficients are statistically significant (all standard errors are less than the obtained estimates, the values of Wald's statistics exceed the critical level, and all error probabilities are less than 0,2).

Let's write down the analytical expression for the constructed multinomial logit model:

$$P(y = 0) = \frac{e^{-1+(-12{,}9)z_3+(25{,}04)z_4}}{1 + e^{-1+(-12{,}9)z_3+(25{,}04)z_4} + e^{3{,}76+(-11{,}2)z_3+(19{,}7)z_4}}$$

$$P(y = 1) = \frac{e^{3{,}76+(-11{,}2)z_3+(19{,}7)z_4}}{1 + e^{-1+(-12{,}9)z_3+(25{,}04)z_4} + e^{3{,}76+(-11{,}2)z_3+(19{,}7)z_4}}$$

$$P(y = 2) = \frac{1}{1 + e^{-1+(-12{,}9)z_3+(25{,}04)z_4} + e^{3{,}76+(-11{,}2)z_3+(19{,}7)z_4}}$$

The resulting expression can be used to assess academic success of students at different values of the factors.

In particular, figures 1 and 2 show the dependence of the studied indicator when changing the factor x_3 (students' grade in computer science) for certain values of the factor x_1 (assessment of the students' presence in the classroom) and the fixed value of the factors x_2 and x_4 ($x_2 = 6$, $x_4 = 3$).

Figures 1 and 2 show an increase in the probability of receiving a better grade and a decrease in the probability of obtaining a bad grade in the subject of information systems in management, given the increasing in the students' grade in computer science. Different values of the students' presence in the classroom have a great influence on the dynamics of increasing the probability of getting a better grade.

We can say that the key to measuring student achievement is assessing student progress by measuring positive changes in the assessment of the students' living conditions and students' presence in the classroom, as well as gradual changes in grade in computer science.

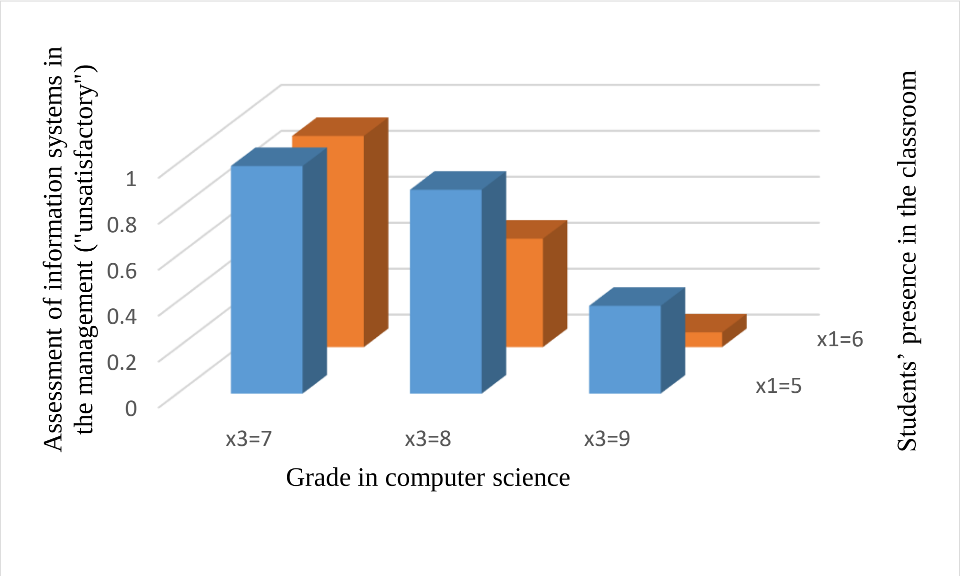

Figure 1: The impact of students' grade in computer science on the assessment of information systems in the management ("unsatisfactory")

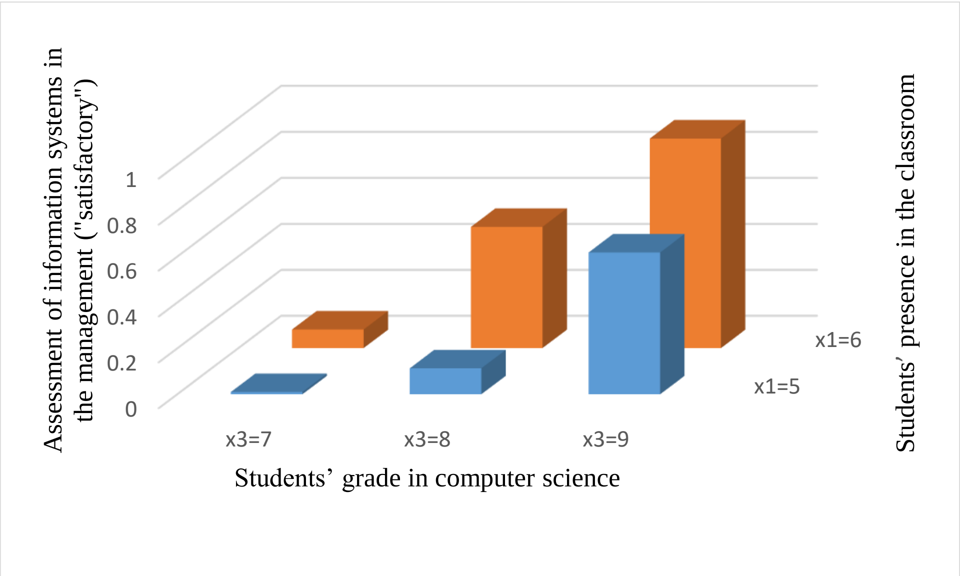

Figure 2: Influence of students' grade in computer science on the assessment of information systems in the management ("satisfactory")

Figures 3 and 4 show the dependence of the studied indicator when changing the factor x_1 (students' presence in the classroom) for certain values of the factor x_2 (students' grade in mathematics) and the fixed value of the factors x_3 and x_4 ($x_3 = 7$, $x_4 = 2$).

Figures 3 and 4 show an increase in the probability of receiving a better grade and a decrease in the probability of obtaining a bad grade in the subject of information systems in management, given the increasing in the students' presence in the classroom. Different values of the students' grade in mathematics have a great influence on the dynamics of increasing the probability of getting a better grade.

We can say that the key to measuring student achievement is assessing student progress by measuring positive changes in the assessment of the students' presence in the classroom and students' living conditions, as well as gradual changes in grade in mathematics.

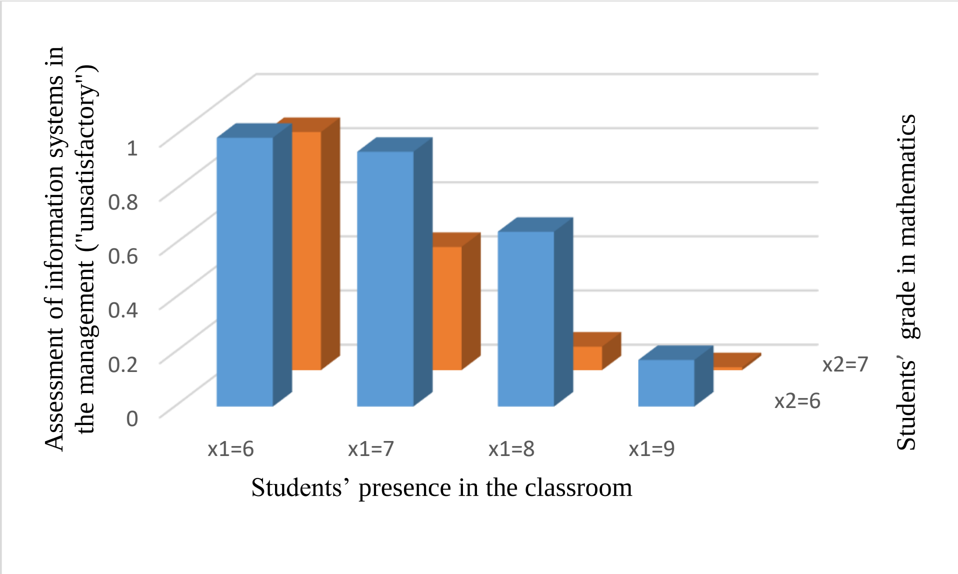

Figure 3: The impact of students' presence in the classroom on the assessment of information systems in the management ("unsatisfactory")

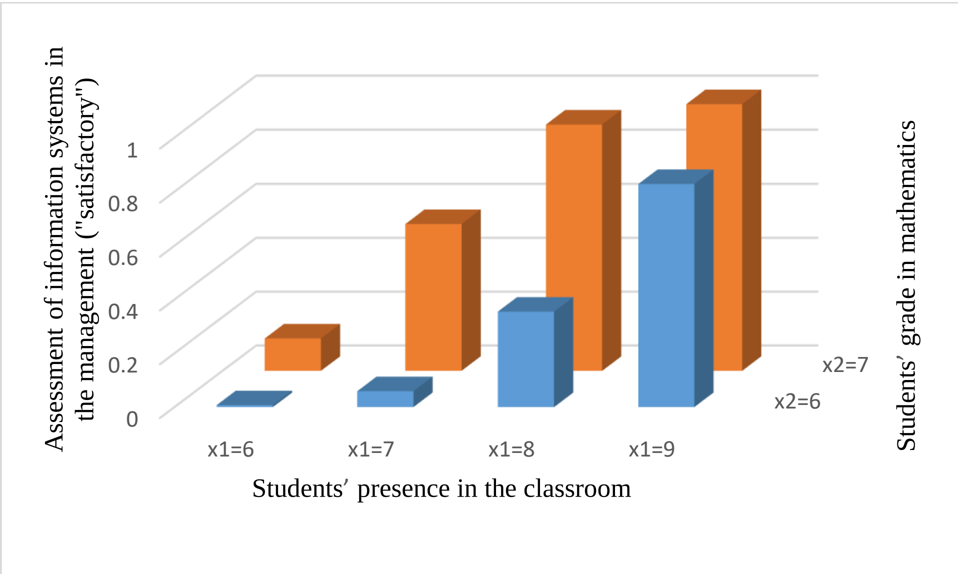

Figure 4: The impact of students' presence in the classroom on the assessment of information systems in the management ("satisfactory")

Figures 5 and 6 show the dependence of the studied indicator when changing the factor x_3 (students' grade in computer science) for certain values of the factor x_2 (students' grade in mathematics) and the fixed values of the factors x_1 and x_4 ($x_1 = 7$, $x_4 = 1$).

Figures 5 and 6 show an increase in the probability of receiving a better grade and a decrease in the probability of obtaining a worse grade in the subject of information systems in management, given the increasing in the students' grade in computer science. Different values of students' grade in mathematics have a great influence on the dynamics of increasing the probability of getting a better grade.

We can say that the key to measuring student achievement is assessing student progress by measuring gradual changes in grade in computer science and grade in mathematics, as well as positive changes in assessment of the students' presence in the classroom and students' living conditions

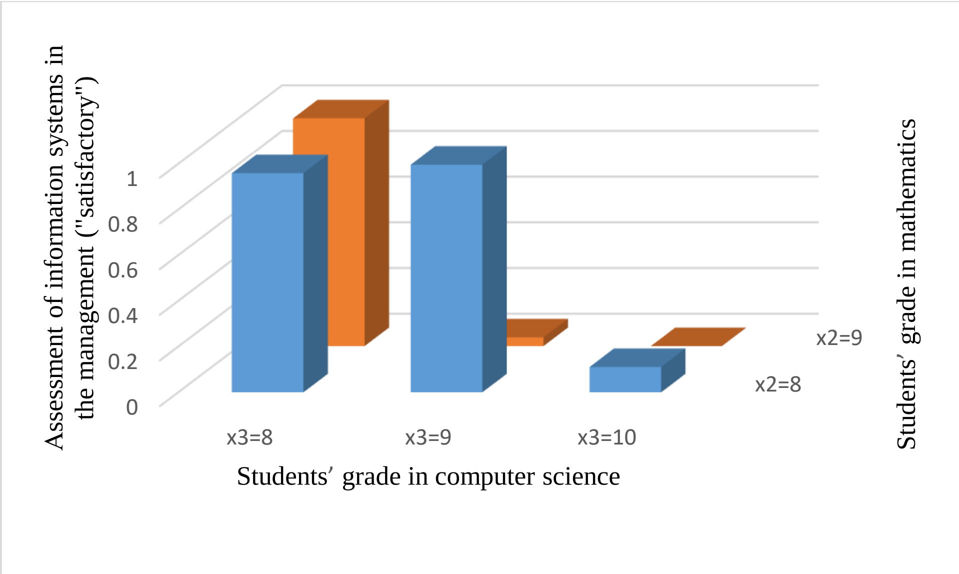

Figure 5: Influence of students' grade in computer science on assessment of information systems in the management ("satisfactory")

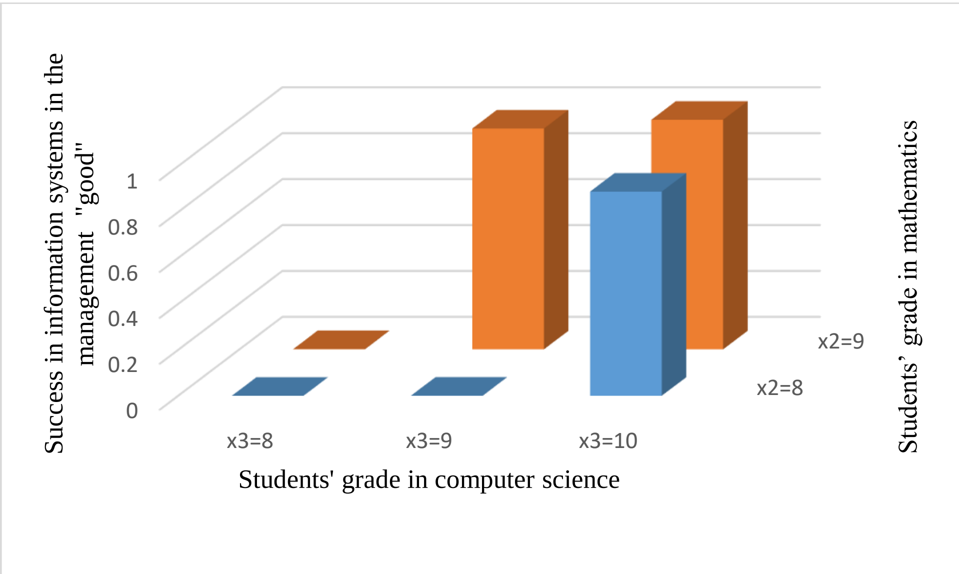

Figure 6: Influence of students' grade in computer science on assessment of information systems in the management ("good")

The goal-oriented approach has a solid foundation for improving the effectiveness of student learning. We can rely on modeling results to develop pedagogical frameworks that help students develop an approach to learning.

Now let's evaluate academic success of students through the developed regression model of an academic success of students. Information for assessing academic success of students is presented in table 3.

Table 3
The value of indicators to assess academic success of students

№	x_1	x_2	x_3	x_4	$P(y=0)$	$P(y=1)$	$P(y=2)$
1	7	10	5	1	0,19147	0,80853	0,00000
2	9	7	8	2	0,00006	0,07106	0,92888
3	6	8	6	2	0,82918	0,17082	0,00000

In the first case, we are likely to receive the grade of "satisfactory". In the second case, we are expecting to receive the grade of "good". It is only in the latter case that a getting the grade of "unsatisfactory" would be most likely.

It is possible to predict individual learning outcomes, to evaluate student actions and differences between actions, to assess factors related to success, to create an individual learning plan through the constructed model. The presented approach has a solid foundation for improving the effectiveness of student learning. We can rely on modeling results to develop pedagogical frameworks that help students develop an approach to learning.

5. Discusion

The multinomial logistic regression model confirms the possibility of combining quantitative (intellectual) and qualitative (behavioral) factors. Our study confirms that student's presence in the classroom, student's grade in mathematics, student's grade in computer science, and student's living conditions affect academic success of students. The study highlighted that academic success of students is driven by variant factors of students' activities. It has been found that the marks for examinations (student's grade in mathematics; student's grade in computer science) significantly affect success. Important quantitative (intellectual) factors that determine the academic success of students are student's grade in mathematics, student's grade in computer science. Important qualitative (behavioral) factors that determine the academic success of students are student's presence in the classroom and student's living conditions.

The current study's conclusion accords with considered research regarding academic success of students. The current study's conclusion accords with considered research regarding academic success of students. Student's presence in the classroom are much likelier determines the main indicator we are researching (Lakkaraju). Student's living conditions are usually strongly affects the productivity of the organization (Oyedeji and other scholars).

The findings indicate that selected factors can significantly affect academic success of students, and creating an educational environment for students, and developing ways to motivate students. This study proved that the directions of the students' activity can be translated into particular approaches to maintain of academic success.

5. Conclusions

The implementation of the regression method to study educational processes in higher education institutions is considered an innovative method. The use of the regression analysis methods allows identifying implicit relationships between elements of the educational process and solve many problems related to the analysis and prediction of academic success of students. It also allows you to build individual educational trajectories. The utilization of these obtained regression relations has a great importance to predict the academic success of students on the basis of statistical information on the control measures of the curriculum. They make it possible to establish a quantifying impacts of disciplines on special disciplines.

This innovative approach will have a greater effect in creating incremental learning behaviors and focusing on the influence of behavioral (non-intellectual) factors. Behavioral factors are difficult to operationalize and suffer from poor prognostic validity. The calculation strategy shows that the mutual influence of intellectual and behavioral factors in students' success can have an impact.

The research methodology can be used to determine the dependence of student learning outcomes on various factors. This is necessary to improve the quality of education in universities. Creative application of regression equations allows carrying out the quality control for each training course, based on individual work with students agreed within teachers of the departments. This research showed to be instrumental in demonstrating practices based on a probit regression model at the level of development of academic success of students for support for the successful completion of their tasks by students, better acquisition of knowledge, acquisition of competencies and profession. The proposed method allows educational institutions to improve the learning process, make learning complete, increase student achievement strategies.

6. Referenses

[1] Z. Kovacic, Early Prediction of Student Success: Mining Students’ Enrolment Data, Proceedings of the Informing Science and Information Technology Education Joint Conference, Cassino, Italy, 19–24 June, 2010, available at: https://proceedings.informingscience.org/InSITE2010/InSITE10p647-665Kovacic873.pdf (accessed 3 January 2022)

[2] H. Lakkaraju, E. Aguiar, C. Shan, D. Miller, N. Bhanpuri, R. Ghani, K. Addison, A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes, Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 21st, 2015, available at: https://dl.acm.org/doi/pdf/10.1145/2783258.2788620 (accessed 21 January 2022)

[3] A. Oyedeji, A. Salami, O. Folorunsho, O. Abolade, Analysis and Prediction of Student Academic Performance Using Machine Learning, Journal of Information Technology and Computer Engineering, 4(1), 2020, pp. 10–15.

[4] S. Alhusban, M. Shatnawi, M. Yasin, I. Hmeidi, Measuring and Enhancing the Performance of Undergraduate Student Using Machine Learning Tools, Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Copenhagen, Denmark, 24–26 August, 2020; pp. 261–265.

[5] M. Pagano, K. Gauvreau, H. Mattie, Logistic Regression. In: Principles of Biostatistics, Chapman and Hall/CRC, 2022

[6] M. Yilmaz, Multinomial logit model, 2018, available at: https://www.researchgate.net/publication/335336335_MULTINOMIAL_LOGIT_MODEL (accessed 21 January 2022)

[7] K. B. Salem, A. B. Abdelaziz, Conduct a multivariate analysis by logistic regression, La Tunisie médicale, 98(7), 2020, pp. 537-542

[8] T. Q. Asenso, H. Zhang, Y. Liang, Pliable lasso for the multinomial logistic regression, Communication in Statistics - Theory and Methods, 2020, available at: https://doi.org/10.1080/03610926.2020.1800041 (accessed 21 January 2022)

[9] C. Ryan, Categorical Data Analysis. In: Data Science with R for Psychologists and Healthcare Professionals, CRC Press, 2021, available at: https://doi.org/10.1201/9781003106845-22 (accessed 21 January 2022)

[10] G. James, D. Witten, T. Hastie, R. Tibshirani, Statistical Learning. In: An Introduction to Statistical Learning, Springer Science+Business Media, LLC, 2021, pp. 15–57, available at: https://doi.org/10.1007/978-1-0716-1418-1_2 (accessed 21 January 2022)

[11] M. R. Abonazel, R. A. Farghali, Liu-type multinomial logistic estimator, Sankhya B, 81(2), 2019, pp. 203-225.

[12] J. H. Kim, Multicollinearity and misleading statistical results, Korean journal of anesthesiology, 72(6), 2019, pp. 558-569.

[13] M. Abonazel, R. A. Farghali, Liu-Type Multinomial Logistic Estimator, Sankhya B: The Indian Journal of Statistics, 81(2/2), 2019, pp. 203-225.

[14] R. Yurynets, Z. Yurynets, M. Kokhan, Econometric analysis of the impact of expert assessments on the business activity in the context of investment and innovation development, CEUR Workshop Proceedings, 2604, 2020, available at: http://ceur-ws.org/Vol-2604/paper47.pdf (accessed 16 January 2022)