

PAPER • OPEN ACCESS

## Modeling the impact of University students research work on the results of their final certification

To cite this article: A Yu Melnykov *et al* 2020 *J. Phys.: Conf. Ser.* **1691** 012187

View the [article online](#) for updates and enhancements.

### You may also like

- [Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence](#)  
Ruth S DeFries, Jessica Fanzo, Pinki Mondal et al.
- [The role of collective action in the cacao sector in enhancing sustainability, market upgrading and agro-biodiversity conservation](#)  
Ximena Rueda, Romaïke Middendorp and Sergio Puerto
- [Social equity in sustainability certification systems for the built environment: understanding concepts, value, and practice implications](#)  
Adam Yeeles, Kim Sosalla-Bahr, Jennifer Ninete et al.

# Modeling the impact of University students research work on the results of their final certification

A Yu Melnykov<sup>1</sup>, N Yu Shevchenko<sup>1</sup>, Ye P Isakova<sup>1</sup> and E Yu Bobkova<sup>2</sup>

<sup>1</sup> Donbas State Engineering Academy, Kramatorsk, 39, Mashinostroiteley blvd, Kramatorsk, Donetsk region, 84301, Ukraine

<sup>2</sup> K.G. Razumovsky Moscow State University of technologies and management (the First Cossack University), 73, Zemlyanoy val str., Moscow, 109004, Russian Federation

E-mail: alexandr@melnikov.in.ua

**Abstract.** The problem of the quality of education is formulated as the central problem of the educational process of the higher education institution. It is emphasized that the final certification is an integral indicator that takes into account all the knowledge and skills acquired during the period of study in various disciplines and other "activities", one of which research work of students (NIRS) is. The task of predicting the influence of students' research activities on the results of their final certification is formulated. Methods of linear multifactor regression and artificial neural networks as a possible mathematical toolkit for predicting are described. It is shown that the best predicting result is provided by the method of artificial neural networks with a perceptron architecture with 8 input factors and two hidden layers with 5 neurons in each. It is indicated that the proposed approach to predicting can be applied when planning the department's activities, for example, when correcting the curriculum of specialties, syllabuses of scientific disciplines, while adjusting the department's management strategy regarding the interaction of students with academic supervisors.

## 1. Introduction

### 1.1. Statement of the problem

Each higher education institution (university) is responsible to the state for the quality of education. Therefore, the problem of quality is central in the educational process of the university. The quality of educational activity is characterized by different components, which include such concepts as: availability of the state standard of higher education and quality of its implementation; quality of the faculty of the university; quality of the learning process organization; quality of the methodological support of the educational process, as well as quality of providers of education or provision of the training of experts in specialties and qualifications determined by the working curricula [1].

The internal system for assessing the quality of education is organized in the form of intermediate and final certification of students. At the same time, it should be taken into account that the final certification is an integral indicator, which is a projection of the knowledge and skills acquired during the training period in different disciplines and other "activities" of students. One such "activity" is the scientific and research work of students (NIRS), which includes the study of specialized disciplines, writing research papers, including course projects aimed at forming skills of identification and description of problem situations etc. [2].



Thus, we can assume that the quality, scale and timeliness of research work directly affects the results of the final certification of students, namely, the average score and average rating of the diploma and the grade that the student receives at the protection of the qualification work. Based on this assumption, having determined the full set of input factors, it is possible to predict the results of the final certification of students based on the results of their intermediate certification and scientific “activities”.

### *1.2. Analysis of recent research and publications*

Quite a lot of research has been devoted to the problem of the impact of research work on the quality of education. In the work [3] the structure and content of students’ research activity, its forms and types, peculiarities of estimating the effectiveness of different types and principles of assessment are analyzed. It is proposed to use a rating system for evaluating NIRS, where the student receives an appropriate rating for each indicator. The study [4] emphasizes the need to identify factors that affect the quality of students’ learning in order to direct the common efforts of all subsystems to achieve a high level of assimilation of knowledge received by students. The authors conducted a mass survey of students and determined that the student’s performance is largely influenced by the personal characteristics of a student.

The authors of [5] investigated a set of 27 factors and selected seven significant ones: 1. average mark of the certificate; 2. attendance; 3. availability of additional education; 4. type of family; 5. number of children in the family; 6. higher education of parents; 7. provision with teaching aids and computer equipment. It is interesting that the categories “Influence of employment in extracurricular time”, “Influence of parents” and “Organization of the educational process” lose their importance at the beginning of the 3rd year.

In the scientific literature, enough attention is paid not only to the analysis of factors that determine the quality of training, but also to methods of predicting academic performance. For example, in the article [6] it was proposed to predict the performance of an individual student in a certain discipline using artificial neural networks. The following indicators were selected as the input factors for the neural network: the average student score in previous related disciplines, the full development of which is a prerequisite; availability of free visits; assessment of residual knowledge testing, conducted before the beginning of studying this discipline; debts in other disciplines. However, with sufficient adequacy (prediction error less than 4%), the proposed model is not adapted for the application in predicting the overall student performance.

The dependence of the quality of education from students’ individual characteristics, teaching staff qualifications and level (“prestige”) of the university is considered in [7]. A polynomial logistic regression analysis was used as a mathematical toolkit for processing questionnaire results.

### *1.3. The aim of the article*

The aim of the article is to study the possibility of predicting the impact of research work of students of a higher educational institution on the results of their final certification by identifying influence factors and constructing an adequate mathematical model.

## **2. Methods of research**

Prediction is the establishment of a functional dependency between dependent and independent variables in order to determine unknown data [8]. When solving the forecasting problem, missing or future values of the target numerical indicators are determined on the basis of historical data. Among the most common prediction methods are linear multifactor regression methods [9] and artificial neural network method [10].

Correlation analysis [8] is used to quantify the relationship of two data sets in order to reduce the number of factors in the construction of analytical dependencies (models). It makes it possible to associate data sets according to their values. The correlation coefficient, which can take a value from -1 to +1 inclusively, indicates the level of communication. The connection between signs (gender Cheddock scale) can be strong, medium, and weak.

When predicting students' performance, we will come out of the assumption that the significance weight of research at the third educational level (master's program) is much higher than at the second (bachelor's).

As input factors, according to our opinion, influence students' overall academic performance, we define: a specialty in which a student is studying in a magistracy; average rating of Bachelor's Diploma (ARB); a mark obtained for the protection of the final qualification work of the bachelor (MB); the total number of student publications of the student (including theses of reports for scientific conferences) articles in co-authorship with the research supervisor (Sprof); number of student publications in specialized ("professional") editions, including articles in co-authorship with the research supervisor (Sprof); last name of the research supervisor (head of the final qualification work); results of the intermediate certification in the discipline "Fundamentals of scientific research" (rating score of the test); results of the intermediate certification in the discipline "Methodology and organization of scientific research" (rating score of the test).

These factors are expected to affect: average score of the master's degree (ASM); average rating of the master's degree (ARM); a mark obtained for the protection of the final qualification work of the master (MM).

Among the disciplines of influence, we select disciplines that directly develop students' ability to carry out scientific and research work (the disciplines "System Analysis" and "Information Systems and Technologies" at Donbas State Engineering Academy (DSEA) are considered). There are "Methodology and organization of scientific research" (MOSR) and "Fundamentals of Scientific Research" (FSR).

At the stage of the pre-prediction analysis of the factors of influence, a correlation analysis was carried out to determine the internal correlation between factors of the same level. The analysis showed a tight connection (0.98) between the average score and the average rating of the master's degree. This is natural, as according to the current "Regulations on the organization of the educational process in DSEA" [11 scholarships for students for the next semester are awarded according to the average rating for the previous semester. Students who are interested in receiving scholarships are persistent in studying all disciplines, regardless of what form the final control has (exam or test). At the same time, the correlation between other close factors, namely, the average rating and a mark obtained for the protection of the final qualifying work, is slightly lower (~ 0.65-0.75). Obviously, a student who already has a permanent job cannot secure a high rating and is not even interested in receiving a scholarship, but high intellectual level allows him to be rated higher when protecting a diploma.

According to the results of the correlation analysis, we will finally operate with eight input and two output factors. Output factors: average rating of the diploma and a mark obtained for the protection of the qualification work.

From a mathematical point of view, the above-mentioned assumptions regarding the dependence of initial and input factors can be formally represented in matrix form:

$$R = WF, \quad (1)$$

where R is matrix of the implementations of the resulting factors of dimension ( $m \times 3$ );

W is matrix of the coefficients that characterize the influence of an input factor on the resulting factor, dimensions ( $3 \times 8$ ); F is matrix of the implementations of the input factors of dimension ( $m \times 8$ ); m is number of students.

Thus, we have the task of predicting the impact of the effectiveness of research on the results of the final certification of graduates of a higher educational institution, which can be carried out by mathematical methods given in the previous section.

Calculations will be carried out in several stages. First, it is necessary to determine the most effective combination of input factors and a more convenient method of calculation (linear multiple regression or artificial neural network). Second, for the chosen method and list of factors, it is necessary to determine such model parameters that will ensure maximum forecasting accuracy [12].

For each case, we will determine the average accuracy of initial factors according to three different approaches:

- average errors in determining the average rating and marks obtained for protection of the qualification work, according to the "100-point" scale;
- average errors in determining initial factors when reducing the data to a “three-point” scale (the number of marks “excellent”, “good” and “satisfactory” is calculated; the percentage of incorrectly determined estimates will be considered an error);
- average errors in determining initial factors in data transforming to the “European” scale (the number of ratings “A”, “B”, “C”, “D” and “E” is calculated; the error is the percentage of incorrectly determined estimates).

### 3. Results and discussion

The studies were carried out on the basis of the results of training students of the specialties “System Analysis” and “Information Systems and Technologies” at DSEA during 2017 – 2019 (see table 1). For the purity of the experiment in this study, the educational process in the selected groups was carried out according to the new standards of higher education published in 2016, and the curricula of the specialties remained practically unchanged over the specified period of time. In addition, only information about students receiving a bachelor's degree on the basis of full secondary education (without an accelerated form of education) was taken into account.

**Table 1.** Students performance.

Spec	Year	Name	ARB	MB	Spub	Sprof	Supervisor	FSR	MOSR	Rating	MM
SA	2018	Student 1	87,16	100	4	1	Shevchenko	100	90	91	99
IST	2018	Student 6	79	91	3	0	Isikova	100	85	80,2	85
SA	2017	Student 13	84,55	97	7	1	Melnykov	92	55	69,2	95
IST	2017	Student 17	71,11	71	2	0	Ivchenkova	55	55	60,67	90

The impact of input factors will be assessed taking into account the following conditions: 1) basic conditions – analysis of the impact of the main learning outcomes at the bachelor’s level (ARB and DB); 2) variable conditions – a combination of other input factors. All calculations were performed in the Deductor Studio Lite environment [13]. Both methods were used: linear multifactor regression (LMFR) and artificial neural network (MLP). The results are summarized in Table 2 and Table 3. It can be seen that both prediction methods demonstrate a fairly accurate result when using 7-8 input factors. Linear multifactor regression provides a more accurate prediction result without using the "Specialty" factor. However, under other equal conditions, the method of artificial neural network is more accurate.

**Table 2.** The average errors in the determination of initial factors for LMFR.

№	ARB	MB	FSR	MOSR	Spub	Sprof	Super-visor	Spec	ARM	MM	Avr	ARK	MMR	Avr	ARKE	MMKE	Avr
1	+	+							0.0346	0.0350	0.0348	0.2500	0.3000	0.2750	0.7500	0.4000	0.5750
2	+	+	+						0.0344	0.0330	0.0337	0.2500	0.3000	0.2750	0.6500	0.3500	0.5000
3	+	+		+					0.0151	0.0302	0.0226	0.1500	0.2500	0.2000	0.4000	0.4000	0.4000
4	+	+	+	+					0.0148	0.0293	0.0220	0.1500	0.2500	0.2000	0.4000	0.3500	0.3750
5	+	+			+				0.0316	0.0307	0.0311	0.3000	0.3500	0.3250	0.7500	0.4500	0.6000
6	+	+			+	+			0.0306	0.0311	0.0308	0.3000	0.3500	0.3250	0.7500	0.4500	0.6000
7	+	+	+		+	+			0.0299	0.0280	0.0290	0.2000	0.3000	0.2500	0.6000	0.4000	0.5000
8	+	+		+	+	+			0.0142	0.0298	0.0220	0.1000	0.3000	0.2000	0.3500	0.5000	0.4250
9	+	+	+	+	+	+			0.0124	0.0279	0.0202	0.1500	0.3000	0.2250	0.4000	0.4000	0.4000
10	+	+	+	+	+	+	+		0.0093	0.0196	0.0144	<b>0.0500</b>	<b>0.1500</b>	<b>0.1000</b>	<b>0.2500</b>	<b>0.2500</b>	<b>0.2500</b>
11	+	+	+	+	+	+	+	+	<b>0.0092</b>	<b>0.0194</b>	<b>0.0143</b>	0.1000	0.2000	0.1500	0.3500	0.3000	0.3250

**Table 3.** The average errors in the determination of the initial factors for MLP.

№	ARB	MB	FSR	MOSR	Spub	Sprof	Super- visor	Spec	ARM	MM	Avr	ARK	MMR	Avr	ARKE	MMKE	Avr
1	+	+							0.0115	0.0120	0.0118	0.1500	0.1000	0.1250	0.3500	0.3000	0.3250
2	+	+	+						0.0194	0.0035	0.0115	0.1500	0.1000	0.1250	0.3000	0.1000	0.2000
3	+	+		+					0.0048	0.0128	0.0088	0.0000	0.1500	0.0750	0.1500	0.1500	0.1500
4	+	+	+	+					0.0012	0.0025	0.0019	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	0.1000	<b>0.0000</b>	0.0500
5	+	+				+			0.0112	0.0025	0.0068	0.1000	0.0500	0.0750	0.2000	0.1500	0.1750
6	+	+				+	+		0.0048	0.0127	0.0088	0.0500	0.0500	0.0500	0.2000	0.1000	0.1500
7	+	+	+			+	+		0.0142	0.0027	0.0084	0.1000	0.0000	0.0500	0.1500	0.0500	0.1000
8	+	+		+	+	+			0.0003	0.0003	0.0003	0.0500	0.0000	0.0250	0.0500	0.0000	0.0250
9	+	+	+	+	+	+			0.0019	0.0057	0.0038	0.0000	0.0500	0.0250	0.0500	0.1000	0.0750
10	+	+	+	+	+	+	+		0.0003	0.0021	0.0012	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>
11	+	+	+	+	+	+	+	+	<b>0.0000</b>	<b>0.0001</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>

For further calculations, we determine the architecture of the neural network. Comparing the average errors of the determination of initial factors for different architectures (except for the already used MLP 7 (8) x5x2), we choose the most optimal one (see Table 4 and Table 5), i.e. one that will provide the minimum error.

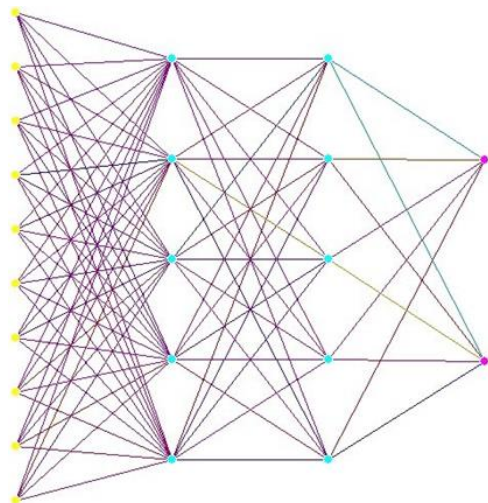
**Table 4.** The average errors of the determination of initial factors for MLP with seven factors.

Network architecture	ARM	MM	Avr	ARK	MMR	Avr	ARKE	MMKE	Avr
MLP 7x5x2	<b>0.0003</b>	<b>0.0021</b>	<b>0.0012</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>
MLP 7x10x2	<b>0.0000</b>	0.0054	0.0027	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	0.0000	0.0500	0.0250
MLP 7x5x5x2	0.0015	<b>0.0008</b>	<b>0.0011</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>
MLP 7x10x10x2	0.0001	0.0143	0.0072	0.0000	0.0500	0.0250	0.0000	0.0500	0.0250

**Table 5.** The average error of the determination of initial factors for MLP with eight factors.

Network architecture	ARM	MM	Avr	ARK	MMR	Avr	ARKE	MMKE	Avr
MLP 7x5x2	0.0000	0.0001	0.0000	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>
MLP 7x10x2	0.0091	0.0099	0.0095	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
MLP 7x5x5x2	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>	<b>0.0000</b>
MLP 7x10x10x2	0.0001	0.0260	0.0130	0.0000	0.0500	0.0250	0.0000	0.0500	0.0250

The minimum value of the average error in determining initial factors is provided by the architecture of the perceptron with two hidden layers and five neurons in each of them (see figure 1).



**Figure 1.** Neural network architecture MLP 8x5x5x2.

The results of the prediction of the average student rating using the MLP 8x5x5x2 model, as well as the actual scores, are given in table 6.

**Table 6.** Forecasting Results.

Student	Rating	Calculation
Student 1	91,00	91,01
...		
Student 6	80,20	80,20
...		
Student 13	69,20	69,20
...		
Student 17	60,67	61,12

#### 4. Conclusions

Studies have shown that the effectiveness of students' research work significantly affects the results of their final certification. The most appropriate mathematical tool for predicting student performance depending on scientific "activities" turned out to be the method of artificial neural networks with perceptron architecture, which consists of eight input factors, two hidden layers with five neurons in each of them. The proposed approach to prediction can be applied when planning the department's activities, for example, when correcting the curricula of specialties, syllabuses of scientific disciplines, and when adjusting the department's managerial strategy regarding the interaction of students with supervisors. As prospects for further researches, it is possible to mark out testing the effectiveness of the proposed approach in predicting the performance of other students of other specialties.

#### References

- [1] Stepanenko I V and Pozdnyakova I R 2010 Criteria for assessing the quality of the educational process at the university *International Journal of Experimental Education* **4** 35-6
- [2] Melnykov A Y and Shevchenko N Y 2014 On the experience of introducing scientific work into the educational process at the Department of ISPR DSEA *Scientific review* **1(2)** 76-80
- [3] Kleshcheva I V 2014 *Evaluation of the effectiveness of research activities of students* (SPb: NRU ITMO) 91
- [4] Mahmutova L R 2018 Factors of influence on student performance at the university *Organization of work with youth* **1** 1182-208
- [5] Shmarihina E S 2018 Study of student performance factors *Bulletin of NSUU* **3** 130-43
- [6] Proshkina E N and Balashova I Y 2018 Analysis and forecasting of student performance based on the radial basic neural network *Engineering: traditions and innovations: materials of the III Intern. scientific conf* (Kazan: Young Scientist) 24-8
- [7] Husain S A and Syed S H 2006 Determinants of education quality: what makes students' perception different? *Open Review of Educational Research* **3(1)** 52-67
- [8] Chubukova I A 2006 *Data Mining: Textbook* (Moscow: Internet University of Information Technologies; BINOM. Knowledge laboratory) 382
- [9] Dubrov A S, Troshin L I and Mhitoryan V S 2003 *Multidimensional Statistical Methods: A Textbook* (Moscow: Finance and statistics) 352
- [10] Hajkin S 2006 *Neural networks: full course* (Moscow: Williams) 1104
- [11] Regulations on the organization of the educational process in the DSEA [http://www.dgma.donetsk.ua/docs/acts/Положення\\_про\\_організацію\\_освітнього\\_2.pdf](http://www.dgma.donetsk.ua/docs/acts/Положення_про_організацію_освітнього_2.pdf)
- [12] Melnykov O Yu 2018 Forecasting of the change in the teacher's contribution to the work of the department using the methods of data mining *Information Technologies and Learning Tools* **1(63)** 266-87
- [13] Deductor is advanced analytics without programming <https://basegroup.ru/deductor/description>