

Intellectual prediction of student performance: opportunities and results

Zolotaryuk A.V.
Financial University
Moscow, Russia
AZolotaryuk@fa.ru

Zavgorodniy V.I.
Financial University
Moscow, Russia
VZavgorodniy@fa.ru

Gorodetskaya O.Yu.
Financial University
Moscow, Russia
OGorodetskaya@fa.ru

Abstract — Intellectual tools for analysis and forecasting are widely used in various fields - economics, medicine, technology, and linguistics. This article examines the possibilities of neural network forecasting of student performance. A general statement of the research object is formulated. Comprehensively considered and classified factors affecting student performance are pre-university, university and psychophysiological ones. The features of the collection of information in Russian universities for intellectual analysis and forecasting are considered. Using the example of the database of the Financial University under the Government of the Russian Federation and additional information obtained by survey about factors, not present in the database, significant factors were determined using the correlation analysis toolset of the IBM SPSS Statistics statistical analysis package, which made it possible to reduce their number almost fourfold and record the regression model in a simpler form. Further research was carried out using a Deductor Studio analytical platform for intellectual processing and knowledge extraction. A multilayer neural network with nine entrance signs and one or two effective ones was built and trained. The effective entrance signs were taken as the results of the first year students taking senior exams. The research results showed that the predicted values of progress do not differ significantly from the actual ones. Consequently, neural network machine study technologies provide intelligent prediction of progress based on an analysis of the preceding factor signs — of both the first and subsequent years. The directions of further research with the use of modern means of machine study are outlined.

Keywords — intelligent prediction; factors affecting progress; regression model; neural networks; progress prediction results

I. INTRODUCTION

Modern intellectual means of analysis and prediction allow, with sufficiently high accuracy, regardless of the subject area, to determine the development trend of the researched variables based on the processing of large arrays of source data and detecting implicit, hidden patterns [1, 2]. There are works on neural network prediction of stock prices and currency rates [3, 4], credit scoring tasks, business strategies and logistics of e-commerce centers [5, 6, 7], features of radio signal propagation [8], pattern, speech and emotion recognition [9, 10, 11, 12], research in medicine [13, 14]. Separate articles, describing the attempts to predict students' progress, were

published [15]. Apparently, the time for a deeper analysis of the possibilities of digitalization in the educational environment, above all, higher education, has come, because the future socio-economic condition of each country depends largely on the skill level of the staff who will come to the economy, other vital areas will implement and deepen current plans and projects, putting forward innovative, highly effective management strategies and the development of various economic sectors, including solutions, found on the basis of modern machine study methods [16; 17].

II. RESEARCH METHODOLOGY

A. General formulation of the problem

Based on detecting, analyzing and evaluating various factors, potentially influencing the results of studies, and choosing innovative digital research and machine study means, to develop a model and perform intelligent neural network prediction of future progress on the example of students of the Financial University under the Government of the Russian Federation (Financial University). Assess the predicted values by comparing them with actual results.

B. Determination of factors affecting the progress of university students

When analyzing the few publications on this problem and, above all, based on the results of the author's research, it was established that all factors S , potentially affecting students' progress in Russian universities can be divided into three groups:

- pre-university factors S_1 ;
- university factors S_2 ;
- psychophysiological factors S_3 .

The group of pre-university factors is quite extensive and can be represented as a hierarchical tree structure (Table 1).

TABLE I. PRE-UNIVERSITY FACTORS

Item	Value
<i>Educational factors S_{11}:</i>	

The presence of the medal "For special achievements in study" on the results of study in the secondary educational institution (school) : • no medal • have a medal	$S_{111} \in Z = [0; 1]$
Average grade certificate of graduation from secondary educational institution	$S_{112} \in R = [3; 5]$
The percentage of excellent grades in the certificate	$S_{113} \in R = [0; 1]$
The percentage of good grades in the certificate	$S_{114} \in R = [0; 1]$
The percentage of satisfactory grades in the certificate	$S_{115} \in R = [0; 1]$
SSE grade in mathematics	$S_{116} \in Z = [0; 100]$
SSE grade in Russian	$S_{117} \in Z = [0; 100]$
Type of secondary educational institution: • comprehensive school • school with deep study of a number of subjects • private individual study school • gymnasium (lyceum) • college	$S_{118} \in Z = [1; 5]$
Socio-regional factors S_{12}:	
Settlement of the educational institution: • capital city (Moscow, St. Petersburg) • a city with a population of over 500 thousand people (except Moscow and St. Petersburg) • a city with a population of 100 to 500 thousand people • a city with a population of up to 100 thousand people • settlement, village, country, etc.	$S_{121} \in Z = [1; 5]$
Sex: • male • female	$S_{122} \in Z = [0; 1]$
Region: • Moscow • Moscow Region • Central FD (except for Moscow and Moscow Region) • Northwest FD • Southern FD • North Caucasus FD • Volga FD • Ural FD • Siberian FD • Far Eastern FD	$S_{123} \in Z = [1; 10]$
Parents: • two parents • one parent • orphan	$S_{124} \in Z = [1; 3]$
Brought up: • as an only child • has one brother or sister • has two or more siblings	$S_{125} \in Z = [1; 3]$
Preuniversity factors S_{13}	
Reason for entering the university: • on the competition according to the USE results and entrance examinations on general reasons • out of competition without passing entrance examinations (winners and prize-winners of All-Russian Olympiads, members of team-winners at international Olympiads, winners of specialized subject Olympiads held by the university, schoolchildren who passed USE with the highest possible grade) • out of competition, taking into account the USE results and passing entrance examinations (disabled, children without parental care, applicants from low-income families, one whose parents is a disabled person of group 1, contract military personnel) • on the competition with a preemptive right of entrance (persons who became disabled as a result of injuries received during the	$S_{131} \in Z = [1; 5]$

performance of military duties, children of persons killed during counterterrorism operations, etc.) • by target admission (according to the contract, concluded with executive authorities, by contract with enterprises and organizations of the defense industry complex)	
Funding of university studies: • budget • commercial	$S_{132} \in Z = [0; 1]$

A number of university factors can also affect the quality of student study (Table 2). Moreover, these factors can play a significant role in all years of study.

TABLE II. UNIVERSITY FACTORS

Item	Values
Educational factors S_{21}	
Methodical skill and experience of the teacher: • high • medium	$S_{211} \in Z = [0; 1]$
Student's motivation: • high • medium • absent	$S_{212} \in Z = [1; 3]$
Information and technological base of the university: • modern • outdated	$S_{213} \in Z = [0; 1]$
The average rating of passing examinations and tests preceding in a university	$S_{214} \in R = [0; 100]$
Socio-economic factors S_{22}	
Student's place of residence: • at home • in dorm • in a rented apartment	$S_{221} \in Z = [1; 3]$
The time spent on the two ways: to the university and back: • up to one hour • up to two hours • up to three hours • more than three hours	$S_{222} \in Z = [1; 4]$
Scholarship: • yes • no	$S_{223} \in Z = [0; 1]$
Parental financial support: • high • medium • absent	$S_{224} \in Z = [1; 3]$
Underwork by student: • none • yes	$S_{225} \in Z = [0; 1]$
Socio-cultural factors S_{23}	
Participation in extracurricular university events: • active, several times a month • medium, not more than once a month • does not participate	$S_{231} \in Z = [1; 3]$
Participation in cultural events: • active, several times a month • medium, not more than once a month • does not participate	$S_{232} \in Z = [1; 3]$
Participation in sports events: • active, several times a month • medium, not more than once a month • does not participate	$S_{233} \in Z = [1; 3]$
Participation in research work: • speaks at student research events • publishes scientific articles • does not participate	$S_{234} \in Z = [1; 3]$
Occupies a leading or elected post: • member of the faculty, institute • the head of the group	$S_{235} \in Z = [1; 3]$

- does not occupy

Personal, psycho-physiological factors are important (Table 3).

TABLE III. PSYCHOPHYSIOLOGICAL FACTORS

Item	Values
Family psychological factors S_{31}	
Dating a girl/boy: • yes • none • recently broke up	$S_{311} \in Z = [1; 3]$
Family status: • single • married	$S_{312} \in Z = [0; 1]$
Children: • yes • no	$S_{313} \in Z = [0; 1]$
Individual psychological factors S_{32}	
Ability to adapt to new conditions: • the ability to quickly build good relations with other people • shyness, unsociability	$S_{321} \in Z = [0; 1]$
Physical health: • good • weak	$S_{322} \in Z = [0; 1]$
Individual features of character: • desire to be a leader • preference to be led	$S_{323} \in Z = [0; 1]$
Ability to high concentration and performance: • yes • none	$S_{324} \in Z = [0; 1]$
The ability to isolate the most important cases and events: • yes • none	$S_{325} \in Z = [0; 1]$

In addition to these, other factors may affect the predicted values of students' progress. For example, it is necessary to take into account whether a student studies in a bachelor's or master's degree, acquires a first higher education or a second, studies on full-time, part-time or abbreviated form of education.

Random events occurring the day before, right before or during an exam or a test may also have a serious impact on student's progress. Unfortunately, such factors cannot be taken into account by means of intellectual digitalization.

C. Features of collecting information for intelligent prediction of students progress

For many years, a rather extensive database of all students has been maintained in Russian universities, including both pre-university information and the results of studies and other activities. Thus, at the Financial University, such a base covers the period from the beginning of this century to the present. As pre-university, the database contains information on factors $S_{116}, S_{117}, S_{131}, S_{132}$. The factor S_{123} is stated for the subjects of Russia (there were 85 of them of March 2019; for many subjects, a few enter the university, or nobody, which makes it impossible to take this factor into account when predicting). Among other significant pre-university factors from the category of pre-university, in the university database there is information about the year of entering the university, faculty, specialty, qualifications, direction and profile of study.

To conduct an intellectual prediction of students' progress, it is necessary to have an array of values of performance signs, in other words, the results (for a financial university, a mark on

a 100-point scale) of students' exams, tests and other types of reporting $Y = \{y_1, y_2, \dots, y_n\}$, in order to train a neural network preliminarily. This information makes up most of the information in the university database, the total amount of which is, in total, gigabytes or millions of lines. However, taking into account that in recent years, our country's universities underwent transformations - larger educational institutions were created by joining smaller ones, new faculties were opened, directions and training profiles, curricula and programs were modified, which was caused by both the needs of the time and the task of integration of education in the Bologna process - all this makes it difficult to directly use the university database to predict a student's progress due to its content heterogeneity and incompleteness of information. Additionally, the accumulation of information about the factors listed in the Tables 1-3 is required.

D. Data preprocessing

After the formation of a representative sample of data, for example, on the basis of a survey of several thousand people, it is possible to build a regression model for obtaining a prediction of the results of passing the exam for students in a certain discipline:

$$Y = f(S_1, S_2, S_3) \quad (1)$$

However, the direct solution of equation (1) is difficult, since the number of factor signs is very large, and all of them affect the performance in different degrees. Preprocessing is required to remove signs, that have little or no effect on the result, as well as those, that are highly correlated with each other. The solution to the preprocessing problem can be found, for example, using the IBM SPSS Statistics analysis package by constructing a pair-correlation matrix. As a result, the number of factor signs can be significantly reduced.

In the conducted research, the most significant factors for first-year students were $S_{111}, S_{112}, S_{116}, S_{117}, S_{118}, S_{121}, S_{122}, S_{123}, S_{131}, S_{132}$. The values of these 10 factor signs, supplemented by effective signs on the passing of advanced exam courses in one or several disciplines at one time by senior students, can be used for neural network prediction of the progress of newly entered students. For the case of prediction in one discipline, the regression model (1) is converted to a simpler equation:

$$y_1 = f(S_{111}, S_{112}, S_{116}, S_{117}, S_{118}, S_{121}, S_{122}, S_{123}, S_{131}, S_{132}) \quad (2)$$

At the same time, the number of observations for training a multilayered neural network will decrease and amount to several hundreds.

Additionally, before the construction and training of the neural network, using the analytical platform of intellectual processing and knowledge extraction Deductor Studio, through which further research will be carried out, we will expose the initial factors and the resultant sign y_1 - the results of first-year students' computer science and information technology exam included in a representative sample. The obtained data (Fig. 1) indicate that there is no correlation of the factors S_{121} (the settlement of the educational institution) with y_1 , which it allows to exclude it from consideration. At the same time, the highest correlation y_1 is observed with the factor S_{112} (the

average grade of the certificate of completion of the secondary educational institution).

Входные поля		Корреляция с выходными полями	
N°	Поле	Y1	
1	S111	0.303	
2	S112	0.714	
3	S116	0.272	
4	S117	0.280	
5	S118	0.046	
6	S121	0.000	
7	S122	-0.207	
8	S123	0.069	
9	S131	-0.168	
10	S132	-0.133	

Fig. 1. Correlation of the resultant sign with factor signs

III. RESULTS OF THE RESEARCH

The 9x6x3x1 multilayer neural network training was implemented for 240 observations with the following processing parameters: teaching set - 95%, test set - 5%, separation of the original set - random, activation function - sigmoid with a slope of 2.00, study algorithm - Resilient Propagation (study in the "offline" mode), completion of study - to achieve 10,000 epochs, take an example recognized, if the error is less than 0.05.

The resulting scatter diagram (Fig. 2) showed that most of the predicted values fit into the selected probability interval.

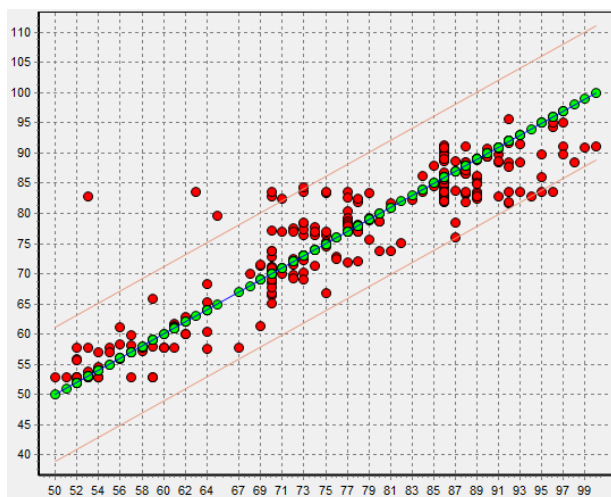


Fig. 2. Scatter diagram

The probability predictive values of Y1_OUT do not significantly differ from the actual Y1 results, and the prediction error Y1_ERR is also insignificant (Fig. 3), which may indicate that the neural network is trained and can be used for intelligent prediction of the results of a new group of students.

Similar researches were conducted to train and predict the results of student progress in other academic disciplines.

Код	S111	S112	S116	S117	S121	S122	S123	S131	S132	Y1	Y1_OUT	Y1_ERR
1	0	4.7	65	95	1	0	1	1	0	91	89.85	0.00053
2	0	4.5	74	92	1	0	1	1	0	85	84.47	0.00011
3	0	4	44	80	1	0	1	3	0	53	57.79	0.00917
4	1	4.9	66	100	3	0	3	1	0	89	83.58	0.01177
5	0	4.3	72	92	2	1	2	1	0	79	75.70	0.00435
6	0	3.8	79	92	1	1	1	1	0	79	78.88	0.00001
7	0	4.4	75	75	2	1	2	1	0	59	65.84	0.01870
8	0	4.8	65	95	1	0	1	1	0	91	88.41	0.00269
9	1	5	55	85	1	0	1	1	0	87	76.06	0.04789
10	1	5	75	75	2	0	3	1	0	86	82.37	0.00528
11	1	4.7	75	75	1	0	1	1	0	91	88.59	0.00232
12	0	4.3	65	85	2	0	2	1	0	71	76.97	0.01428
13	0	4	78	81	3	1	3	3	0	56	58.33	0.00218
14	0	4.3	52	71	1	0	1	1	1	69	71.37	0.00224
15	0	4.1	70	87	2	0	3	1	0	72	71.62	0.00006
16	0	3.9	56	76	1	0	1	1	1	52	55.88	0.00603
17	0	4	73	84	2	1	2	1	0	69	71.46	0.00242
18	0	4.5	64	76	1	0	1	1	1	89	85.92	0.00380
19	0	4.3	77	82	2	0	1	1	0	76	72.45	0.00503
20	0	4.3	54	72	1	0	1	1	1	82	75.02	0.01947

Fig. 3. Fragment of the table with the parameters of actual and predictive values

Using a trained neural network, it is possible, by setting the values of the input parameters for a new student, to obtain a probabilistic predictive mark for passing the corresponding exam (Fig. 4).

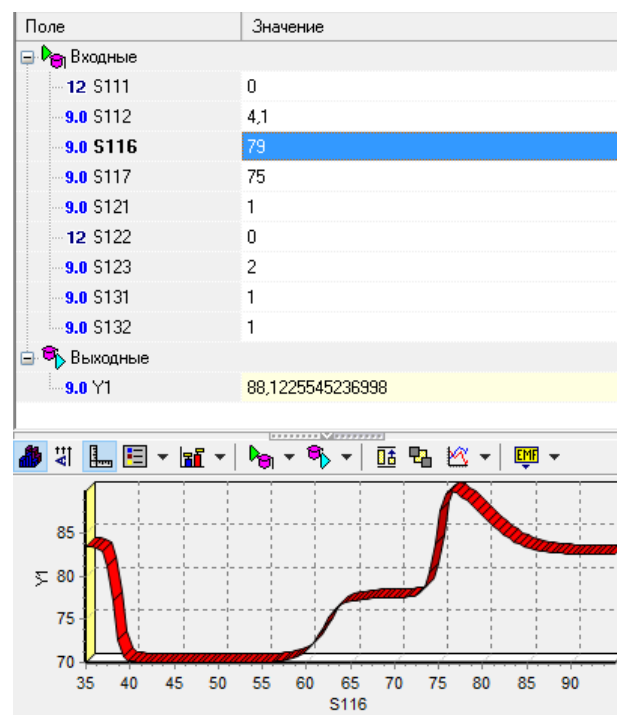


Fig. 4. An example of representation of the predicted value of the student's progress with a schedule of their USE grades on the math exam

The predictions, obtained this way, were subsequently compared with actual data after the session. And in this case, the prediction error mostly did not exceed 7–10%, which fits into the statistical error. Certain surges in one direction or another can be explained by unaccounted random factors, which does not exclude the practical significance of the performed researches.

During the research, it was also established that when predicting the results of students passing the subsequent tests and examinations, university and psychophysical

factors S_2 и S_3 should be taken into account, since, in this case, they have more influence.

IV. CONCLUSIONS (INFERENCE)

During the research, it was established, that intellectual means for analysis and prediction are widely used in various fields - economics, medicine, technology, and linguistics. It is proposed to use this toolset in the educational process.

The general statement of the problem is formulated and the possibilities for predicting students' progress are considered. Comprehensively analyzed and classified factors, affecting progress, - pre-university (educational, socio-regional and pre-university), university (educational, socio-economic and socio-cultural) and psychophysiological (family-psychological and individual-psychological). It was noted that other factors, including random events, may affect students' progress.

The features of the collection of information for the intellectual prediction of student' progress are analyzed. The general characteristics of the database on students of Russian educational institutions are considered on the example of the Financial University under the Government of the Russian Federation. The need for additional survey for students to collect information about the missing factors is noted.

The general regression model for obtaining the prediction of the results of students passing the test or exam is given. The issues of preprocessing input factors are considered to reduce the number of signs by building a pair-correlation matrix using the IBM SPSS Statistics statistical analysis package. The most significant factors, affecting the progress of first-year students are highlighted.

Using the analytical platform for intellectual processing and knowledge extraction, Deductor Studio, the architecture is selected and a multilayer neural network with nine input signs, two hidden layers and one or two effective signs, which were taken as the results of the first year students, is investigated on a representative sample. The results confirmed the proficiency of the neural network

The trained neural network was used to predict the performance of the first year students. A subsequent comparison of the predicted values with the actual ones showed that the errors are minor, not exceeding the statistical error.

To predict the progress of senior students, it is recommended to take into account university and psychophysical factors, which influence increases significantly.

Thus, the technology of neural network machine study provides the possibility of intelligent prediction of student progress on the basis of an analysis of previous factor signs.

V. DISCUSSION OF RESULTS

Preliminary and current research results were reported and discussed at scientific and technical seminars and conferences of the Financial University and received a positive assessment. It was noted that the prediction of student progress at the stage of the selection committee's work, if the corresponding situation is changed, can improve the quality of the students entering the university. In the future, during the study process, the data can be used both by the teaching staff for calculation

in their pedagogical activities and methods of teaching disciplines, as well as by students who, in case of low prediction and perseverance of character and purposefulness, can show better results at the session.

During the discussion, recommendations were made on the use of other modern methods of machine study based on processing of large amounts of information for predicting other modern methods.

References

- [1] Jürgen Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, 2015, pp. 85-117.
- [2] Mehdi Zekriyapanah Gashti, Rouhollah Habibey, "Training Feed-forward Neural Networks using Asexual Reproduction Optimization (ARO) Algorithm", *International journal of neural networks and advanced applications*, vol. 5, 2018, pp. 13-16.
- [3] Sneha Jain, Dr. Roopam Gupta, Dr. Asmita A. Moghe, "Stock Price Prediction on Daily Stock Data using Deep Neural Networks", *International Journal of Neural Networks and Advanced Applications*, vol. 5, 2018, pp. 45-54.
- [4] Shahmelikyan T.A., Kesiyan G.A., Urtenov M.H. "Using correlation analysis with neural network forecasting stock prices", *Fundamental research*, No. 4, 2013, pp. 224 -228.
- [5] Volkova E.S., Gisin V.B., Solov'ev V.I., "Data Mining Techniques: Modern Approaches to Application in Credit Scoring", *Digest Finance*, 2017, vol. 22, iss. 4, pp. 200-412.
- [6] Trifonov Yu.V., Shiryayeva Yu.S., Gromnitskiy V.S., "Formation of enterprise activity strategies using neural networks", *Creative Economy*, vol. 12, No. 6, 2018, pp. 805-812.
- [7] Baiyu Chen, Biying Wang, "Location Selection of Logistics Center in e-Commerce Network Environments", *American Journal of Neural Networks and Applications*, vol. 3, No. 4, 2017, pp. 40-48.
- [8] Kriti Priya Gupta, Madhu Jain, "Performance Analysis of Cellular Radio System Using Artificial Neural Networks", *American Journal of Neural Networks and Applications*, vol. 3, iss. 1, 2017, pp. 5-13.
- [9] Verkholyak O.V., Kaya H., Karpov A.A., "Modeling Short-Term and Long-Term Dependencies of the Speech Signal for Paralinguistic Emotion Classification", *Trudy SPIIRAN – SPIIRAS Proceeding*, vol. 18, No. 1, 2019, pp. 30-52.
- [10] Golovchiner M.N., Dylykov A.R., "The use of neural networks in the task of identifying the dominant personal characteristics of different categories of respondents", *Vestnik of Tomsk State University: Management, Computer engineering and Informatics*, No. 3 (36), 2016, pp. 91-96.
- [11] Kipyatkova I., Karpov A. A "Study of Neural Network Russian Language Models for Automatic Continuous Speech Recognition Systems", *Automation and Remote Control*, vol. 78, No. 5, 2017, pp. 858–867.
- [12] Markovnikov N.M., Kipyatkova I.S. "An Analytic Survey of End-to-End Speech Recognition Systems", *Trudy SPIIRAN – SPIIRAS Proceeding*, 2018, iss. 3(58), pp. 77-110.
- [13] Yasnitsky L.N., "Neural network – a tool for gaining new knowledge: achievements, problems, prospects", *Neurocomputers: development, application*, No. 5, 2015, pp. 48-56.
- [14] Berebin M.A., Pashkov S.V., "Experience of application of artificial neural networks for the purposes of differential diagnosis and prognosis of mental adaptation", *Uzhno-Yralsk State University Journal of Computer technology, Control, Radio Electronics*, No. 14(69), 2006, pp.41-45.
- [15] Pol'din O.V., "Forecasting the academic achievement in the university by the results of USE", *Applied econometrics*, No. 1(21), 2011, pp. 56–69.
- [16] Neubig G. et al. "DyNet: The Dynamic Neural Network Toolkit", 2017. URL: <https://arxiv.org/pdf/1701.03980.pdf>
- [17] Zolotaryuk A.V., "Information technology priorities of higher education in the age of economy digitalization", *New Information Technologies in Education: Proceedings of the 19th International Scientific and Research Conference "New Information Technologies in Education" (Using IC technologies for educational purposes and for human resource development in digital economy) January 29-30, 2019, Edited by D.Chistov, Part 2. – Moscow. LLC "IC Publishing" in 2019, pp. 287-290.*