**Model for determining directions of school graduates based on machine learning algorithms**

Oybek Mallayev a, Jaloliddin Aliyev, Jamoliddin Gazatov

Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

a) Corresponding author: [jaloliddinaliyev0@gmail.com](mailto:jaloliddinaliyev0@gmail.com)

**Abstract**. The purpose of this article is to determine the orientation of schoolchildren in terms of grades in subjects and soft skills using machine learning methods. The article examined the difficulties of constructing a sigmoid function using multivariate linear regression, and also digitized grades obtained in selected subjects in the field of student education over 10 years and their various parameters, reasons and student capabilities. Using these numbers, a training data set was created. As a result, a classification of subjects studied by schoolchildren over 10 years and their assessments was developed. Neural network architectures, modules, the most commonly used activation functions in machine learning algorithms, training methods and methods for constructing linear and logistic regression, disadvantages and opportunities are analyzed. Ways to simplify the gradient descent function for multivariate linear regression by vector calculation have been studied. Because there are many variables involved in this type of linear regression, vector calculations have proven to be more convenient. Methods for parallel calculation of gradient descent processes using vector calculations are also considered. In particular, the addition of training data table columns, transposition of coefficients - AT, vectorized representation of a linear function, hyperparameters for gradient descent (learning rate, number of steps) were defined.

**INTRODUCTION**

This year, a comprehensive analysis of the local education system by experts revealed the following shortcomings in the higher education system:

Lack of integration between theory and practice in higher education. Inefficient organization of students' practical training in industrial enterprises resulted in most graduates needing to re-learn their profession and specialization after employment, instead of graduating as fully prepared specialists.

The mechanism for controlling the quality of education does not meet modern requirements.

Shortage of qualified pedagogical and administrative staff in educational institutions.

Insufficiently effective cooperation with foreign educational institutions.

Inadequate participation of economic sectors in shaping orders for future personnel training, developing qualification requirements for graduates, and ensuring the quality of training specialists needed by the industry.

Disconnection between higher education, science, and production, with a lack of integration.

Other similar deficiencies were identified.

Work is underway to develop mechanisms to address these shortcomings. However, there is another side to the issue.

It's a fact that most students enrolling in higher education find themselves in a field that doesn't match their aptitude or interests. Even with the most modern teaching methods and practical training mechanisms, the quality of graduates remains unimproved.

Based on the essence of this problem, this article proposes an algorithm for proactively identifying highly talented and motivated students interested in their chosen field of study using machine learning algorithms, an artificial intelligence module, and software tools.

Developing an artificial intelligence module to address this issue presents several major challenges, namely parameters requiring considerable time investment:

Studying school student assessment mechanisms.

Analyzing student grades in different subjects across various fields of study.

Identifying students' soft skills in different fields of study.

Forming a list of essential subjects for each field of study.

Generating tables of grades for each student in these essential subjects.

Developing tables that consider age psychology, talents, and achievements.

Developing tables of grades in subjects over several years to ensure the proposed artificial intelligence module produces realistic results.

Gathering this data, which requires a significant amount of time, and identifying its various parameters are crucial for the proposed module.

**Table 1.** Training data

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№** | Academic years | | | | | | | | | | Activity | Achievements | **RESULTS** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 0 | 0 | 1 |
| 2 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 1 | 0 | 1 |
| 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 0 | 1 |
| 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 0 | 1 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 0 | 1 |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 0 | 1 |
| 7 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 1 | 1 |
| 8 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 1 |
| 9 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 1 |
| 10 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 1 |
| 11 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 1 |
| 12 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 0 | 1 |
| 13 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 1 | 0 | 1 |
| 14 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 2 | 0 | 1 |
| 15 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 0 | 1 |
| 16 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 1 |
| 17 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 0 | 1 |
| 18 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 1 | 1 |
| 19 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 2 | 1 |
| 20 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 3 | 1 |
| 21 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 4 | 1 |
| 22 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 0 | 5 | 1 |
| 23 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 1 |
| 24 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 5 | 0 |
| 25 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 0 |
| 26 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 5 | 0 |
| 27 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | 0 |
| 28 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 1 | 0 |
| 29 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 2 | 0 |
| 30 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 3 | 0 |
| 31 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 4 | 0 |
| 32 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 0 | 0 |

Students' knowledge level and comprehension are evaluated through ongoing and interim assessments. Ongoing assessment, in the form of quizzes, tests, or assignments, regularly monitors students' knowledge, skills, and abilities. Interim assessment evaluates students' knowledge, skills, and abilities at the end of each quarter and upon completion of relevant sections of the curriculum. This takes the form of written exams or tests. The materials used in these assessments are prepared by the subject teacher.

**Essential Subjects and Development Parameters for Engineering Majors**

Students aiming for engineering programs should ideally have strong grades in the following subjects:

Mathematics

Physics

For each student, a table of grades obtained over 10 years was compiled for these two subjects (Table 1). Grade change parameters (variants) over the years were identified, resulting in large data tables. In short, this constitutes the input and training data for machine learning algorithms. Increasing the amount of this multi-parameter training data improves the accuracy of the module and reduces errors. As an example, a table of mathematics grades was compiled and supplemented with "activity" and "achievements" parameters. These parameters provide crucial values for the field of study.

The proposed machine learning model was trained using a matrix filled with data from Table 1.

This table shows a single student's mathematics grades over 10 years. The "Activity," "Achievements," and "Results" columns were added based on the criteria mentioned above to train the artificial intelligence module and minimize errors, thus reducing the uncertainty in determining the student's suitability for the field. The proposed module required a minimum training data table (Table 1) of 32 rows for decision-making; the actual table consists of 384 rows. It is enriched with various versions of the student's grades in one subject over 10 years. If the input data is not present in the training data table, the program adds it to the training data table after the result is displayed.

The supervised learning method was used to build the artificial intelligence module proposed in this article.

A crucial aspect of machine learning is the creation of algorithms to teach computers data and allowing the computer to learn independently. Analyzing datasets, selecting an appropriate algorithm, using it to train the data, and analyzing the results are the fundamental elements of machine learning [10]. This field includes the following major areas:

Supervised Learning: In this method, data created for training purposes is provided to the system, and the system learns from this data. This training method is used for many purposes, including predicting learning outcomes and using other applications. This type of training is currently being used to solve many problems. The computer program is trained using training data. Supervised learning is further divided into:

1. Regression

2. Classification

Regression training was selected to build a model for identifying the educational paths of graduating students.

Unsupervised Learning: In this method, data is provided to the system, but their value is not specified. The system analyzes, learns from, and modifies the data. The goal is to train the model based on unlabeled data.

Semi-supervised Learning: This method combines supervised and unsupervised learning. This allows the system to learn from a portion of labeled data and a portion of unlabeled data.

Reinforcement Learning: In this method, the system learns through trial and error, determining future actions and modifying itself to achieve better results.

Linear regression, chosen for the model, is one of the most basic methods in machine learning, used for predicting continuous values.

Deep Learning Technologies

Recently, deep learning technologies have been expanding their capabilities across all fields, being widely used to solve various complex problems. This is aided by increased computing power, large datasets, and the ability to use cloud storage [4].

Neural networks are used in education for:

Complex calculation processes.

Improving the probability of accurate prediction.

Implementing quality, accuracy, and online monitoring of class attendance.

Fast information retrieval and finding high-quality data.

Without activation functions, the output signal would be a linear function, a first-order polynomial. While linear equations are easy to solve, they have limited complexity and therefore have a reduced ability to learn complex functional mappings from data. Thus, without an activation function, a neural network would be a limited linear regression model. The most popular activation functions include:

* **sigmoid**  The function takes a real number as input and transforms it into a value between 0 and 1. Specifically, large negative numbers map to 0, and large positive numbers map to 1
* **Tanh** function  It squashes a real number into the range [-1, 1]. Its output has a zero mean. This can be seen as a tuned sigmoid, the hyperbolic tangent (tanh) function..
* **Softmax** The function produces a result that varies between 0 and 1, and the sum of probabilities equals 1.
* **ReLU (Rectified linear unit)** U (𝑥) = max(0, 𝑥) It evaluates the function. In other words, the activation simply has a zero threshold.
* **ELU (Exponential Linear Units)**  **x** Here, the negative component is modeled using exponential features.

Deep learning is particularly well-suited for complex identification tasks such as face recognition, machine translation, and speech recognition [5][6]. The performance of a neural network is highly dependent on parameter selection. There is no single universal algorithm for determining these parameters; therefore, optimal values are typically found through trial and error. Experimentation with multiple variations using different parameter values is necessary to identify the best settings.

**Mathematical Basis of the Model: Linear Function Formula for the Model**:

 (1)

When performing linear regression, it's necessary to find the best-fitting regression line that represents the data. When calculated using equation (1), XT is obtained. In this case, the predicted value obtained using this constant should be as close as possible to the actual value given in the dataset. That is, the difference between the predicted value and the actual value should be minimal. This difference is called the error. To determine the error, the mean squared error (MSE) function is used.

 (2)

Reducing the error can be achieved by updating the regression coefficients. Therefore, optimization is performed using (2). The result of the module involves the steps of calculating a multiple linear regression for the dependence on several columns of the training data table (Table 1).

To accomplish this, certain parameters/notations were introduced.:

m – Number of data points (number of rows); n– Number of variables (number of columns); x(i) – i row variables; –i – the value of the J - variable in the row; X – Matrix representing variables in the dataset.

aj  – j- coefficient;

A – a matrix or vector representing all coefficients. It measured. (n x 1);

y(i) – i- the original value of the projected character in the row in the dataset;

– is a unit Matrix or vector representing all y(i) s progzoned. Its size (1 x m);

f(x) – forecast calculation formula; In particular, in the data set given in Table 1:

m= 32, that is, the number of data sets (number of rows) is 32;

n = 12, the number of variables (number of columns) is 12. In this, variables are made up of academic years (10 columns), activity, achievement columns;

x(1) – [5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 0, 0, 1] that is, the set of values of the columns of academic years (10 columns), Activity, achievements listed in Row 1;

– 0, i.e. the value (activity) of the 11th variable in Row 1;

y(1) – 1, i.e. the original value of the predictive character in the dataset in line 1;

– (learning rate);

– epoch – the number of iterations over the entire dataset during gradient descent

Applying multiple linear regression to the dataset in Table 1, formula (1) takes the following form:

 (3)

Gradient descent for multiple linear regression is made more efficient through vectorized computation [9]. Because this type of linear regression involves many variables, vectorized calculations offer significant computational advantages. Therefore, modifications are made to equation (3) to leverage the efficiency of vector operations:

 (4)

Applying equation 4 to a specific row of the dataset yields the following equation:

 (5)

5- the formula was implemented for each of the rows of the data table, in which - Predicted value of side 1, - m- the projected value of the series is. The elements of this formula are expressed using a matrix. Matrix of projections:

 (6)

Formula 5 results in a matrix of size (m x 1). The matrix of variables takes the following form:

 (7)

This results in a matrix of size (m x (n+1)), where 'm' is the number of rows (observations) and (n+1) is the number of variables (including the intercept). The following formula represents the coefficient matrix:

 (8)

For vector multiplication, the transpose of vector A (swapping rows and columns) is used, resulting in the following expression:

 (9)

This results in a matrix of size ((n+1) x 1), i.e., a vector. The next step involves multiplying the X matrix by the AT vector.

 (10)

The formula for multivariate linear regression, expressed using matrices as shown in Formula 10, can be written concisely as:

 (11)

In Formula 11, ŷ is a vector of size (m x 1), X is a matrix of size (m x (n + 1)), and AT is a vector of size ((n + 1) x 1)..

To find the coefficients that minimize the cost function using the gradient descent function, the following form of Formula 2 was derived to make the determination and calculation of the cost function for multivariate linear regression more convenient:

 (12)

Since the main goal of gradient descent is to minimize the cost function, the new coefficient does not alter the overall objective.

 (13)

The cost function was determined using Formula 13. To minimize it, we use gradient descent formulas. As is known from single-variable regression, to minimize a function, it is necessary to find the coefficients that lead to the minimum value of this function. If the cost function is expressed using coefficients, it takes the following form:

 (14)

The cost function in Formula 14 involves (n+1) coefficients. These can be written as:

 (15)

To find the minimum value of the cost function, it is necessary to find the optimal values of the coefficients given in Formula 15. For this, the derivative of the cost function with respect to each coefficient is taken. Using the obtained derivatives, we derive the gradient descent formula for multivariate linear regression (16).

In Formula 16, initially, random numbers (the learning rate) are assigned to the values of the coefficients; then, the operations are performed. These operations are repeated (until the end of an epoch) until the optimal values of the coefficients are found. For this purpose, the cost function is calculated at each step of an epoch. The values of the coefficients in these operations are updated in parallel.

 (16)

The model passes the given data through several layers. Initially, multivariate linear regression is calculated because the values in the "Activity" and "Achievements" columns of the training data table (Table 1) are greater than 5. The goal of linear regression is to find a continuous target variable, while logistic regression predicts discrete (categorical) target variables. Since the given training dataset belongs to a dataset with discrete values, a sigmoid function was used in the proposed model.

The sigmoid function is also called the logistic function. Its main function is to determine the value of a given number in the range (0,1). In machine learning, the sigmoid function is used to determine the probability of the predicted value. Its formula is given in Formula 17.

 (17)

Here, z is the input variable to the sigmoid function. The application of the sigmoid function to logistic regression is as follows:

 (18)

Here, the matrix multiplication (X ) yields continuous, i.e., linear, values. To obtain discrete values, a sigmoid function must be applied.

(19)

The complete formula for the sigmoid function in logistic regression is as follows:

 (20)

A shorter representation of the sigmoid function is:

 (21)

In formula 21, Z=XAT.

**CONCLUSION**

This study created a training dataset by analyzing the subjects studied and grades received by secondary school students over 10 years, examining their influence on the choice of educational path. A multivariate linear regression function was used and analyzed, employing this training dataset as input values. Analyses showed that the chosen cost function and gradient descent process in building the proposed machine learning model provided the ability to predict values closest to the given values. The correctness of the multivariate linear regression formula relative to the given values was proven. A sigmoid function is used to determine the probability of the predicted value. The algorithm constructed using the model and the software tool created using the Python programming language determined, based on grades in mathematics and physics over 10 years, student activity, and achievements, which future educational path would allow graduating students to achieve the highest success.

**REFERENCES**

1. H. Zaynidinov, J. Nurmurodov, and S. Qobilov, Application of machine learning methods for signal processing in piecewise‑polynomial bases, Proc. 9th IEEE Int. Conf. Inf. Technol. Nanotechnol. (ITNT), (2023).

2. H. Zaynidinov, M. Singh, U. S. Tiwary, and D. Singh, Lecture Notes in Computer Science, 13741 LNCS, v–vi (2023).

3. H. Zaynidinov and O. Mallayev, Parallel algorithm for calculating the learning processes of an artificial neural network, AIP Conf. Proc. 2647, 050006 (2022). https://doi.org/10.1063/5.0104178

4. O. U. Mallayev, Parallel algorithm for digital processing of medical images based on the parallelism paradigm, Innov. Technol. Sci. Educ. Conf., 2(11).

5. H. Zaynidinov, O. Mallayev, and M. Kuchkarov, Parallel algorithm for modeling temperature fields using the spline method, Proc. IEEE IEMTRONICS, 9422645 (2021).

6. H. Zaynidinov, O. Mallayev, and J. Nurmurodov, Parallel algorithm for constructing a cubic spline on multi‑core processors in a cluster, Proc. 14th IEEE Int. Conf. Appl. Inf. Commun. Technol. (AICT), 9368680 (2020).

7. H. N. Zaynidinov, O. U. Mallaev, and B. B. Anvarjonov, A parallel algorithm for finding the human face in the image, IOP Conf. Ser.: Mater. Sci. Eng. 862, 052004 (2020).

8. Yu. A. Volchek, O. N. Shishko, O. S. Spiridonova, and T. V. Mokhort, Position of artificial neural network models in medical expert systems, Juvenis Scientia, no. 9 (2017).

9. H. Zaynidinov, S. Ibragimov, G. Tojiboyev, and J. Nurmurodov, Efficiency of parallelization of Haar fast transform algorithm in dual‑core DSPs, Proc. 8th Int. Conf. Comput. Commun. Eng. (ICCCE), 7–12, 9467190 (2021).

10. H. Zayniddinov, B. Rakhimov, G. Khalikova, and A. Saidov, Review and analysis of computer vision algorithms, AIP Conf. Proc. 2789, 050022 (2023).