МРНТИ 14.35.09
УДК 004.378

https://doi.org/10.51889/1134.2022.25.74.017

STUDY OF IMPACT SIGNS ON THE QUALITY OF EDUCATION BY MACHINE LEARNING ALGORITHMS

A.K. Bazarbai[1], A.K. Berdaly[1], Z.M. Abdiakhmetova[1], V. Jotsov[2]*

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]University of Library Studies and Information Technologies, Sofia City, Bulgaria
[*]e-mail: arassatbazarbay@gmail.com

Abstract

Currently, the rapid development of information flow, the expansion of new digital services affects the quality of education of students. In the field of education, many scientists have studied in their works the identification of signs that affect student progress. In the research work, a database is created, including the physiological and psychological characteristics of students. Predictions are made based on the database with machine learning algorithms. Linear regression, Support vector machine, Random Forest methods are used to determine the best algorithm. Algorithms are evaluated using Regression Evaluation metrics. The result of the study provides the best algorithm and necessary indicators for learning progress. The main goal of the article is to study the signs of impact on the quality of education with the help of machine learning algorithms and write and analyze the results in the Python programming language.

Keywords: machine learning algorithm, python programming language, support vector method, random forest, dataset, linear regression, features.

Аңдатпа

МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІМЕН БІЛІМ САПАСЫНА ӘСЕР ЕТУ БЕЛГІЛЕРІН ЗЕРТТЕУ

А.К. Базарбай[1], А.К. Бердалы[1], З.М. Абдиахметова[1], В. Йоцов[2]
[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан,
[2]Кітапхана ісі және ақпараттық технологиялар университеті , София қ., Болгария

Қазіргі уақытта ақпарат ағынының қарқынды дамуы, жаңа цифрлық қызметтердің кеңеюі студенттердің білім сапасына әсер етуде. Білім беру саласында көптеген ғалымдар өз еңбектерінде оқушылардың үлгеріміне әсер ететін белгілерді анықтауды зерттеген. Зерттеу жұмысында оқушылардың физиологиялық және психологиялық ерекшеліктерін қамтитын мәліметтер қоры жасалады. Болжамдар мәліметтер базасы негізінде машиналық оқыту алгоритмдерімен жасалады. Ең жақсы алгоритмді анықтау үшін сызықтық регрессия, Қолдау векторлық машинасы, Random Forest әдістері қолданылады. Алгоритмдер Регрессияны бағалау метрикасының көмегімен бағаланады. Зерттеу нәтижесі оқу үлгерімінің ең жақсы алгоритмі мен қажетті көрсеткіштерін береді. Мақаланың негізгі мақсаты – машиналық оқыту алгоритмдерінің көмегімен білім сапасына әсер ету белгілерін зерттеу және нәтижелерді Python бағдарламалау тілінде жазу және талдау.

Түйін сөздер: машиналық оқыту алгоритмі, Python программалау тілі, тірек векторлық әдісі, кездейсоқ орман, деректер жинағы, сызықтық регрессия, белгілер**.**

Аннотация

А.К. Базарбай[1], А.К. Бердалы[1], З.М. Абдиахметова[1], В. Йоцов[2]
[1]Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан,
[2]Университет библиотечного дела и информационных технологий, г. София, Болгария

ИССЛЕДОВАНИЕ ПРИЗНАКОВ ВЛИЯНИЯ АЛГОРМИТМОВ МАШИННОГО ОБУЧЕНИЯ НА КАЧЕСТВО ОБРАЗОВАНИЯ

В настоящее время стремительное развитие информационных потоков, распространение новых цифровых сервисов влияет на качество образования студентов. В области образования многие ученые изучали в своих работах выявление признаков, влияющих на успеваемость учащихся. В исследовательской работе создается база данных, включающая физиологические и психологические особенности студентов. Прогнозы делаются на основе базы данных с алгоритмами машинного обучения. Линейная регрессия, метод опорных векторов, методы случайного леса используются для определения наилучшего алгоритма. Алгоритмы оцениваются с использованием показателей регрессионной оценки. Результат исследования обеспечивает наилучший алгоритм и необходимые показатели успеваемости. Основная цель статьи – изучить признаки влияния на качество

образования с помощью алгоритмов машинного обучения и записать и проанализировать результаты на языке программирования Python.

Ключевые слова: алгоритм машинного обучения, язык программирования python, метод опорных векторов, случайный лес, набор данных, линейная регрессия, признаки.

Introduction

Currently, the trend of digitalization of society is going very fast. This, in turn, simplifies many tasks encountered in everyday life. Of course, this phenomenon will find a place in the field of education and will have an impact on the development of this field. Lydia Sandra, Ford Lumban Gaol, and Tokuro Matsuo [1] reviewed factors that may affect student performance and technologies that help predict student performance.

To date, lesson plans have been drawn up in general terms, so they are the same for all students. However, students have different types of learning abilities, so the same lesson plan may not be ideal for all students. Imagine a scenario where a student is able to learn quickly with visuals/shapes/diagrams, but is provided with text-based learning material - the material may be difficult for the student to absorb. Before ML and machine learning, there was no practical way to detect this and come up with a possible solution. As a result, this puts a lot of pressure on the student and sometimes leads to failure even though the student may have had good potential. If the material were presented only in a different way, then the student could easily understand and assimilate it.

The use of artificial intelligence is a great solution for this situation. Individual lesson plans can potentially lead to better learning because technology can evaluate student data and determine the best practices in which students can learn. This will also determine the best subject mapping based on student interests.

The study addressed many issues, one of which was to identify important characteristics that could be used to predict student performance. The results indicated that internal evaluation and summative evaluation were the most consistent traits used to predict academic performance. In addition, other important traits such as personal and internal evaluation, previous records, extracurricular activities and social attributes were identified. One of the main goals of this work was to study the use of machine learning algorithms to predict the progress of students. The study was conducted in mid-May 2021 and the data collected is from the IEEE Access and Science Direct databases. Data collection was performed according to standard database search rules: exclusion and inclusion criteria, as well as search results. The work includes logistic regression, Bayesian method, K-nearest neighbor (KNN), regression tree, random forest, decision tree, long-short-term memory (LSTM), support vector method (SVM), multi-layer perceptron neural network (MLP), algorithms such as artificial neural network (ANN) are used. Artificial neural network (ANN), support vector method (SVM), decision tree, Bayesian method, logistic regression methods are algorithms that have shown good results [2].

Another study [3] on this topic states that due to the large amount of data in the database in the field of education, it has become difficult to predict the progress of students. The main purpose of the research work is to provide an overview of artificial intelligence systems used to predict academic learning. The research paper presents an automated evaluation system for evaluating the progress of students and analyzing their achievements. Here, the author uses a tree-like algorithm to accurately predict student progress. The data clustering method was used to analyze a large set of students' databases. It can be seen that this method speeds up the search process and gives accurate classification results. A new model of teaching was proposed using information about the student obtained during college registration. The final data set is provided as input so that machine learning algorithms can apply and predict student performance. In the study, 13 algorithms of 5 categories were selected from machine learning algorithms: Bayesian, SVM, MLP, IBK, linear regression and tree-type algorithms. As a result of the research, using Binomial Logistic Regression gave 97.06% accuracy, Decision Tree 88.24%, Entropy 91.18%, K-Nearest Neighbor showed 93.72% accuracy. Among the machine learning algorithms, Binomial logistic regression gave the best results.

The following research paper [4] argues that the economic success of any country largely depends on the availability of higher education and that this is one of the main concerns of any Government. According to Hussein Altabrawee, American student loan debt has increased because many students are unable to complete higher education on time. As for machine learning methods, fully connected artificial neural network, Bayesian, decision tree and logistic regression were used in the work to build a machine learning model. The ROC (receiver operating characteristic) index was used to compare the accuracy of the four models. The dataset used to build the models was collected from liberal arts college students during the 2015 and 2016 academic years using a student survey and test booklet. The dataset contains information on 161 students. The activities of this study include creating a student dataset, collecting data, pre-processing the data, building and

evaluating four machine learning models, finding the best model, and analyzing the results. The composition of the data set is wide-ranging: from the learner's age and gender, to the extent to which family members help with the student's lessons. The effect of the above-mentioned four models on the data set was expressed by the ROC index. Artificial Neural Network has a value of 0.807, Bayesian 0.697, Decision Tree 0.762 and Logistic Regression 0.767 and the most efficient model is found to be Artificial Neural Network [5]. Many other works have been carried out in this direction, and a brief overview of those works is given according to several properties (Table 1).

Table 1. Review of Research Papers

Article	*Features*	*Data set size*	*Machine learning algorithm*	*Best Algorithm*	*Features*
Meier et al, 2015 [4]	*Prices*	*700*	*The proposed new algorithm, KNN, linear regression, logistic regression, SVM*	*A new proposed algorithm*	*The number of signs is small*
Guleria et al, 2014 [5]	*Progress in class, participation in class, performance of assignments, laboratory work, session progress*	*120*	*Decision tree*	*Decision tree*	*Number and algorithm of small signs*
Xu et al, 2017 [6]	*Evaluation, work experience*	*1169*	*Linear regression, logistic regression, RF, kNN, proposed progressive prediction algorithm*	*Proposed progressive prediction algorithm*	*The number of signs is small*
Arsad et al , 2013 [7]	*Prices*	*896*	*ANN*	*ANN*	*Number and algorithm of small signs*
Li et al, 2013 [8]	*Prices*	*72*	*PCA*	*PCA*	*Number and algorithm of small signs*
Gray et al, 2014 [9]	*Abilities, personality, motivation of learning strategy*	*914*	*NB, DT, logistic regression,SVM, ANN, KNN SVM, KNN, NB*	*SVM, KNN, NB*	*The number of signs is small*
Buniyamin et al, 2016 [10]	*Prices*	*391*	*Neuro-fuzzy classification*	*Neuro-fuzzy classification*	*The number of signs is small*
Alharbi et al, 2016 [11]	*student demographics, overall performance, student Modules*	*898 for training 1789 for testing*	*Logistic regression, ANN, DL, BN, DA, DT and ensemble approach*	*There are no overall winners*	*The best model is not defined*
Livieris et al, 2012 [12]	*Prices*	*279*	*ANN, DT, NB, rule learning, SVM*	*ANN, SVM*	*Small number of signs*
Hamsa et al, 2016 [13]	*Internal Grades, Sessional Grades and Drop Score*	*168*	*Fuzzy Genetic Algorithm and DT*	*The FGA model is less rigorous than the DT*	*Few data sets*
Sarker et al, 2014 [14]	*Personal and demographic information, student satisfaction and integration*	*149*	*ANN, logistic regression*	*logistic regression*	*Fewer datasets and algorithms*
Huang et al, 2011 [15]	*Grade point average and grades*	*239*	*Linear regression, ANN, SVM.*	*SVM*	*Small number of signs*

Studying the aforementioned works, we noticed that they lacked the scope of data sets, the methods used during data collection, the lack of predictive features, and the lack of machine learning algorithms.

Machine learning can be used to predict student learning outcomes. The system is accurate, sensitive and specific in terms of using machine learning algorithms to predict student performance in the early stages of learning [5]. They mainly perform the following tasks:

– Automatic collection of exam scores.

– Exam performance data Implicit variables are then discovered based on these data, such as prior knowledge, talent, and diligence of the student.

Based on this information, a prediction model is created to predict student learning outcomes.

2. Materials and methodology

Educational software must be very accurate as it affects the learning process of students. They are also used for assessment and tutoring of students. Hence, they must go through an intensive testing process before implementation. Software testing verifies that the software conforms to the attributes of the system, as well as its ability to achieve its intended goals. As software complexity grows, the testing process becomes more intense [6]. The learning elements are software metrics and specification, control graph, call graph execution data, test case failure report, and coverage data.

Software testing includes the following steps:

– Analysis of the problem area and its corresponding data sets.

– Algorithm analysis.

– Analysis of implementation options. An automated testing process is used to reduce the cost of testing and the time required to complete it.

Intelligent learning environment refers to systems that use both supervised and unsupervised learning. It is very useful for representing student learning in terms of knowledge, metacognitive abilities, and learning behavior in order to predict the student's future behavior. These systems are usually based on statistical pattern recognition. These models are achieved by collecting, processing, examining and testing data. This can be done by manually labeling the data and then applying supervised learning algorithms to identify behavior. To further reduce the implementation time, unsupervised learning algorithms such as k means clustering to identify common learning behavior can be used, and supervised learning is implemented to actually build the user model based on the identified patterns.

2.1. Participants

Data sets from the Internet source (https://archive.ics.uci.edu/ml/datasets/student+performance) were used in this research. This dataset includes 395 (187 girls, 208 boys) students. The participants were students from the Portuguese schools named Gabriel Pereira and Musinho da Silveira, without neurological or mental disorders [6].

2.2. Selection of signs

Data attributes include student grades, demographic, social, and school characteristics and are collected through school reports and surveys. Indicators include age and gender of learners, parents' work and healthy daily diet. The number of signs is equal to 33. The signs were followed by properties that may have been overlooked given the shortcomings of the aforementioned research, but are of significant predictive value (Figure 1).

2.3. Work performed

Machine learning algorithms trained by three different teachers were applied to the sorted dataset. In particular, Linear regression, Support vector machine, Random Forest methods. The textual value data was converted to numerical value using the LabelEncoder module of the Python programming language. The data was split into two because the machine learning algorithms were trained by the teacher. To train the preliminary model, a training data set and a test data set for prediction were separated [7].

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10
GP	M	16	U	LE3	T	4	3	services	other	...	5	4	2	1	2	5	10	15	15	15
GP	M	16	U	LE3	T	2	2	other	other	...	4	4	4	1	1	3	0	12	12	11
GP	F	17	U	GT3	A	4	4	other	teacher	...	4	1	4	1	1	1	6	6	5	6
GP	M	15	U	LE3	A	3	2	services	other	...	4	2	2	1	1	1	0	16	18	19
GP	M	15	U	GT3	T	3	4	other	other	...	5	5	1	1	1	5	0	14	15	15
GP	F	15	U	GT3	T	4	4	teacher	health	...	3	3	3	1	2	2	0	10	8	9
GP	F	15	U	GT3	T	2	1	services	other	...	5	2	2	1	1	4	4	10	12	12
GP	M	15	U	LE3	T	4	4	health	services	...	4	3	3	1	3	5	2	14	14	14
GP	M	15	U	GT3	T	4	3	teacher	other	...	5	4	3	1	2	3	2	10	10	11
GP	M	15	U	GT3	A	2	2	other	other	...	4	5	2	1	1	3	0	14	16	16

Figure 1. Data set

Linear regression is a method that determines the model of dependence of variable x with a linear dependence function on one or more other variables (factors, regressors, independent variables). Capturing the dependence between input and output variables in a linear model distinguishes the linear regression method from other methods. If we consider the dependence between a single input and a single output variable, the regression is simple and the following regression equation is defined for it: y= ax + b. The work was done using the Linear Regression module built into the Python programming language. We worked with split data using fit() and predict() methods. The final results were obtained in the form of an array [8].

Support vector machine is a teacher-trained machine learning model with training algorithms that analyze data and identify patterns. The method of support vectors for both classification problems, also used for regression analysis. The peculiarity of the method is the determination of planes dividing all points into classes [9]. The Python programming environment helped to simplify the work. Because the Support Vector method is a built-in method in this environment. After invoking the module, the linear type of kernel that divides the points into classes of the method is selected. Using the Predict () method, the variables were predicted and the result was obtained in the form of an array.

Random Forest is a Machine Learning method used to solve regression and classification problems. It uses ensemble learning, which is a method that combines many classifiers to solve complex problems. In the work, the RandomForestRegressor library was launched and the necessary parameters were set. Data were predicted using the fit() and predict() methods of this module. The final result is achieved in the form of an array [10].

Results

Most efficient classifiers in a range of disciplines are support vector machines (SVM). It can also be used to diagnose and treat cardiovascular disease. With a modest margin of separation between the two objects, this method works effectively in high-dimensional situations. Aside from samples, it works well with a wide range of other dimensions as well another drawback of this approach is that it requires a lot of time and memory to run. As a result, the performance metrics may suffer while dealing with noisy and confusing data [11]. SVM, despite its flaws, offers a number of valuable applications because of its appropriate and cost-effective categorization. Linear regression, Support vector method and Random forest algorithms were applied to the collected data set. These algorithms have achieved good results for prediction purposes. As mentioned above, the result is obtained in the form of arrays. Figure 2 shows the results obtained by the Linear Regression method.

```
array([ 5.92431866, 10.70981935, 10.48604256,  6.97682276,  1.33269023,
       -3.16271817,  5.80856325,  4.8255278 ,  5.04927279,  7.02938248,
        9.2116912 ,  7.87801775,  7.88734034, 11.18645587,  3.94567677,
       -2.82710306, 13.61026047,  6.24229375,  5.2546867 ,  7.59215563,
       12.31582624, 12.23445724, 10.67683712,  6.35567804,  5.63038122,
        6.25061206, 11.99087005,  9.68436036,  4.72345842,  6.13400196,
        7.16816436,  2.63963535, 13.67916046, 11.50850488,  7.66692063,
        8.73483111,  6.62385271,  2.83258597, -1.78469759,  5.60751638,
       13.53935825, 12.63807143,  5.67415779,  8.17924455,  4.20132426,
       10.94647203,  7.81055369,  4.86244401,  8.51589707, 10.03441079,
```

Figure 2. Linear regression result

Using the linear kernel of the support vector method, the indicators in Figure 3 were obtained.

```
array([ 6.07381508, 10.97065736, 10.95551508,  8.03815575,  1.88848695,
       -3.02123005,  6.09314178,  5.18012788,  5.68262987,  7.94833358,
        9.15958054,  8.06596321,  7.99897873, 11.83887521,  4.98010759,
       -2.47341315, 13.24048775,  6.23696593,  5.92906032,  7.95649432,
       12.23170433, 12.09152301, 10.95297858,  6.53102662,  5.88605232,
        6.76356503, 12.03043622,  9.8716055 ,  5.7365891 ,  6.15057343,
        7.13935305,  2.95078734, 13.17860609, 12.08603449,  7.1990561 ,
        8.0697712 ,  7.15049114,  3.72652255, -2.41525341,  6.01225867,
```

Figure 3. The result of the method of support vectors

The Random Forest method using ensemble training gave significantly more accurate results compared to other methods. Figure 4 shows the result of the Random Forest method.

```
array([ 2.29357143, 10.94483333, 10.84483333,  7.48116667,  1.07058333,
        1.20836905,  6.54      ,  4.56033333,  5.59226299,  7.48066667,
        8.61242063,  7.59442857,  7.88792857, 12.81577381,  4.66458333,
        1.21445238, 13.31678571,  6.48366667,  4.45307143,  7.555     ,
       12.42875   , 12.33802381, 10.915     ,  6.13904762,  6.17107143,
        6.63064286, 12.47845238, 10.32211905,  5.60965476,  6.30640476,
        6.38765476,  3.29444048, 13.39319048, 12.64203571,  7.62233333,
        7.5055    ,  6.77416667,  3.86552381,  0.669     ,  6.13240476,
```

Figure 4. The result of the random forest method

In the study, the performance of the algorithms was evaluated using Regression evaluation indicators.

$$\frac{1}{m}\sum_{1}^{m}(y_i - \hat{y}_i)^2 \quad (1)$$

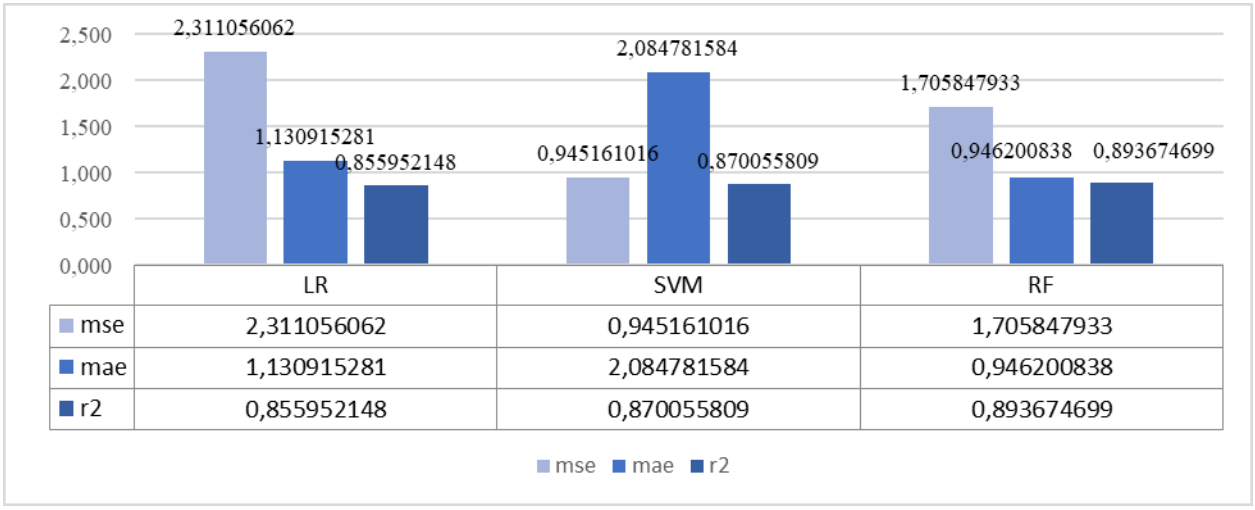
$$\frac{1}{m}\sum_{1}^{m}|(y_i - \hat{y}_i)| \quad (2)$$

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \quad (3)$$

Mean squared error (1), mean absolute error (2) and R-squared (3) indicators are calculated due to the main expected and predicted variables. The work of the algorithms was calculated and compared with the following formulas (1,2,3) [13].

According to the results of these evaluation indicators, the mean square error of the linear regression method is 2.311, the mean absolute error is 1.131, R square is 0.856, the mean square error of the support vector method is 0.945, the mean absolute error is 2.085, the R square is 0.870, the mean square error of the random forest method is 1.706, the mean absolute error 0.946, R square was equal to 0.894.

The results are shown in Figure 5.



	LR	SVM	RF
mse	2,311056062	0,945161016	1,705847933
mae	1,130915281	2,084781584	0,946200838
r2	0,855952148	0,870055809	0,893674699

Figure 5. Analysis

Beside the limitations, the research has proposed a set of features to enhance effective e-learning in order to build and improve the quality of the education system by current popular techniques. At the same time, we proposed research on facilities to improve the performance of training institutions based on successful application of artificial intelligence [14].

In the classification problem, the characteristics of the data always affect the classification results. Depending on the purpose of the problems, they propose different factors that affect different systems. Therefore, the number and value of the attributes will be one of the factors determining the model results. The question of how to best complete a course is always posed to each student. It depends on many influencing factors, but this information is not always fully provided by the student.

The random forest method has shown to be the most suitable method with its indicators. Figure 6 shows the signs that affect the quality of a student's knowledge horizontally and vertically.

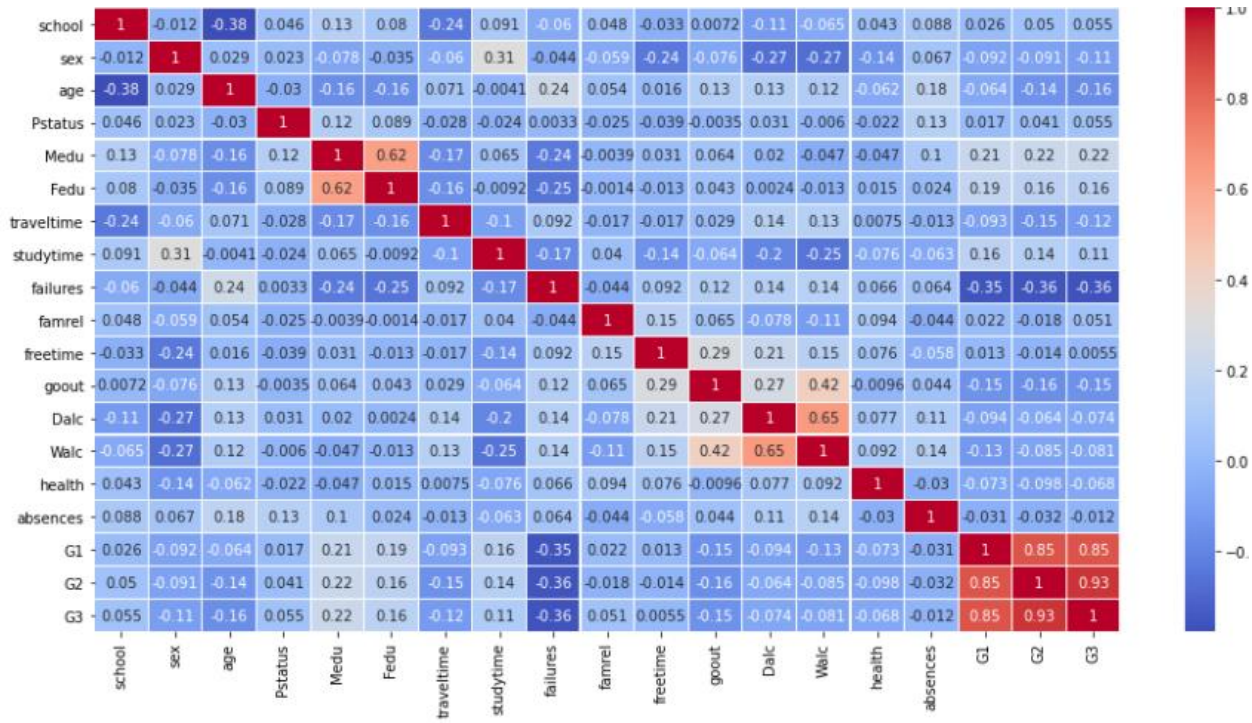


Figure 6. Correlation table

This correlation table shows the relationship of these features. To evaluate the connection, we used a scale from 0 to 1, which means that 0 indicates a low connection and 1 indicates a high connection. The table shows that the classes of the student's father and mother, healthy eating on weekdays and weekends of the week and grades for the previous quarter are closely related.

Discussion

This article looked at regression techniques for machine learning. The progress of the students takes an important place in the educational process. Therefore, in the research work, prediction and assessment of student performance was carried out. The prediction was made using a database created on the basis of student questionnaires and school reports. The database contains various symbols. Characteristics such as the student's age and gender, daily walking time and access to the Internet were included. It is worth noting that it is a feature of the signs. Since signs directly affect the final result in forecasting, it is worth paying attention to their selection. Improving our research, in the future it is planned to add such important indicators as the "index of use of digital tools". Of course, these predictions were made using machine learning algorithms. Linear regression, support vector method and random forest methods have achieved good prediction results. The basis for this is the regression evaluation indicators used to evaluate the performance of the algorithms. According to the research, the profession of the parents influences their child's education. If we take a look at real life, we will witness that a student whose parents work in the field of education deepens in the field of education and reaches scientific degrees [15].

In today's competitive world, an institute must be able to forecast student performance, classify people based on their talents, and seek to improve that performance in future exams in order to be successful. Prior to beginning a course, students should be instructed to focus their attention on one single topic area. With the use of studies like this, an institute may be able to reduce its failure rates. This study predicts students' success in a specific course based on their previous achievement in similar courses. Machine learning, an approach for discovering hidden patterns in massive volumes of existing data, is involved. These patterns may be useful in the future for study and predictions.

Significance of the study - In the era of artificial intelligence, machine learning facilitates the individual teaching and learning process to improve the effectiveness of educational applications. The efficiency and effectiveness of the education system can be improved through dynamic adaptive learning and teaching strategies. In modern pedagogical systems, there are only a few teaching systems that are dynamic and able to meet the individual needs of students.

The availability of these systems should be increased by including agents and learning objects in educational applications. Such intelligent learning systems should be adaptive, capable of learning and dynamic. Many educational technologies are projects available either as standalone learning systems or as web-based learning tools. All of these projects use methods such as multimedia interaction, learning models, and asynchronous learning. For the architectural design of pedagogical information, the necessary integrated approach is provided.

Conclusion

In conclusion, the results showed that the used algorithms have a high predictive ability. The research paper explained the obvious importance of signs for prediction. Learner performance predictions powered by machine learning algorithms support teachers in making early predictions of student performance and making decisions about improving student performance. In addition, machine learning algorithms can be used to both predict high-achieving students and identify students at risk of failure, so that learners at risk of failure can receive early and timely additional support.

Of course, these are the useful aspects of machine learning that we know, and we think that in the future we will recognize many unknown aspects through further research and analysis. In the future, it is planned to improve the models and take into account the signs that are invisible in forecasting, but can have a significant impact on the model's performance. In order to increase the accuracy of prediction, the addition of new signs is being considered. The obtained results allow us to determine the effective factors for improving the educational process and determine the perspective of this research work.

References:

1 *Lidia Sandra, Ford Lumbangaol, Tokuro Matsuo Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review, Vol. 10, Issue 4, pp. 1919-1927, 2021.*

2 *J. Dhilipan, N.Vijayalakshmi, S.Suriya, Arockiya Christopher Prediction of Students Performance using Machine learning, pp. 0-8, 2020.*

3 *Hussein Altabrawee, Osama Ali, Amir Qaisar Predicting Students' Performance Using Machine Learning Techniques, Vol.(27), pp. 194-205, No.(1): 2019.*

4 Y. Meier, J. Xu, O. Atan and M. Van Der Schaar, "Predicting grades," IEEE Trans. Signal Process, vol. 64, no. 4, pp. 959–972, 2016.

5 P. Guleria, N. Thakur, and M. Sood, "Predicting student performance using decision tree classifiers and information gain," Proc. 2014 3rd Int. Conf. Parallel, Distrib. Grid Comput. PDGC 2014, pp. 126–129, 2015.

6 J. Xu, K. H. Moon and M. Van Der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE J. Sel. Top. Signal Process., vol. 11, no. 5, pp. 742–753, 2017.

7 P. M. Arsad, N. Buniyamin and J. L. A. Manan, "A neural network students' performance prediction model (NNSPPM)," 2013 IEEE Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2013, no. July 2006, pp. 26–27, 2013.

8 K. F. Li, D. Rusk and F. Song, "Predicting student academic performance," Proc. – 2013 7th Int. Conf. Complex, Intelligent, Softw. Intensive Syst. CISIS 2013, pp. 27–33, 2013.

9 G. Gray, C. McGuinness and P. Owende, "An application of classification models to predict learner progression in tertiary education," in Souvenir of the 2014 IEEE International Advance Computing Conference, IACC 2014, 2014.

10 N. Buniyamin, U. Bin Mat and P. M. Arshad, "Educational data mining for prediction and classification of engineering students achievement," 2015 IEEE 7th Int. Conf. Eng. Educ. ICEED 2015, pp. 49–53, 2016.

11 Z. Alharbi, J. Cornford, L. Dolder and B. De La Iglesia, "Using data mining techniques to predict students at risk of poor performance," Proc. 2016 SAI Comput. Conf. SAI 2016, pp. 523–531, 2016.

12 E. Livieris, K. Drakopoulou and P. Pintelas, "Predicting students' performance using artificial neural networks," Proc. 8th Pan-Hellenic Conf. "Information Commun. Technol. Educ., pp. 28–30, 2012.

13 H. Hamsa, S. Indiradevi and J. J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," Procedia Technol., pp. 46-18, 2016.

14 F. Sarker, T. Tiropanis and H. C. Davis, "Linked data, data mining and external open data for better prediction of at-risk students," in Proceedings - 2014 International Conference on Control, Decision and Information Technologies, CoDIT 2014, pp. 18-21, 2014.

15 S. Huang and N. Fang, "Work in progress: Early prediction of students' academic performance in an introductory engineering course through different mathematical modeling techniques," Proc. - Front. Educ. Conf. FIE, vol. 1, pp. 3–4, 2012.