МРНТИ 14.35.09 https://doi.org/10.47751/skpu.1937.v41i3.1

DETERMINATION OF STUDENTS' ACADEMIC PERFORMANCE USING MACHINE LEARNING METHODS

Nakhipova Venera Ismailovna[1], Suleymenova Laura Askarbekkyzy[2], Adylbekova Elvira Tulepbergenovna[3].

[1]Doctoral student, [2]Candidate of Tehnical Science, [3]Candidate of Pedagogical Science.

[1, 2, 3] Zhanibekov University, Shymkent/Kazakhstan.

[1]ORCID: 0000-0002-4248-8943, e-mail: nakhipovavenera@mail.ru

[2]ORCID: 0000-0002-4579-1622, e-mail: laurasuleimenova7174@gmail.com

[3]ORCID: 0000-0003-1471-0137, e-mail: adylbekova_elvir@mail.ru

Abstract

The research paper examines important aspects of using approaches to learning analytics (LA) and educational data mining (EDM), which can be used to develop educational programs and track progress. It has been shown that the integration of learning analytics into a learning management system (LMS) can improve the effectiveness of the educational process by predicting and preventing problems faced by students with regard to academic performance. The process of analyzing and predicting academic performance data using machine learning methods, including linear regression, with the determination of the ratio of academic performance in a discipline and GPA points is considered. The result of the study shows the importance of choosing the right functions to improve the accuracy of forecasting using machine learning models, and also provides input data to improve the educational program. The process of creating predictive models for assessing student academic performance by collecting and preparing data using machine learning algorithms is considered. A clear description of the use of models to predict the future academic performance of students is given, and special attention is paid to the interpretation of the results obtained to identify the main factors affecting academic performance. This work contributes to the development of the field of learning analytics and the fields of educational data production, offering practical recommendations for curriculum management and improving the quality of data-based education.

Keywords: machine learning, artificial intelligence, LMS, learning analytics, intellectual analysis of educational data, predictive analytics, linear regression.

	Accepted: 30 September 2024.
	Corr. Author: Suleymenova L.A., e-mail: laurasuleimenova7174@gmail.com
For citation:	Nakhipova V.I., Suleymenova L.A., Adylbekova E.T. (2024). Determination of students' academic performance using machine learning methods. *Ilim* 41(3). 5-20.

Information about financing

The work was carried out with the financial support of a grant from the Ministry of Education and Science of the Republic of Kazakhstan (grant “integration of machine learning to predict learning strategies in LMS in the formation of professional competencies of students”).

Introduction

Currently, machine learnig technologies have proven their importance for the realization of learning outcomes for students all over the world, including in the field of education. Education has become more than just the teaching of texts or the requirement to memorize manuscripts, but an attempt to measure the goals and results of the learning process in and out of the classroom. Over time, teaching methods have become a dynamic part of the income and expenses of the educational process. In addition, these practices have become an important part, playing a major role in promoting the components of the learning system, updating the foundations of the curriculum, increasing its effectiveness and ingenuity. These components are used in the process of planning, implementing, evaluating, monitoring and developing goals (Rossman and Chance, 2014: 211-221).

Machine learning has become a new milestone in higher education. Currently, machine learning is gaining significant interest from the education sector in the development of artificial intelligence technologies.

Artificial intelligence and machine learning are often talked about at the same time. Machine Learning (ML) is an artificial intelligence method for classifying and profiling, both supervised and unsupervised. For example, predict the probability that a student will leave the course or be accepted into the program, or identify topics in written assignments. Terms such as artificial intelligence, deep learning, and machine learning are now widely used in education and by its professionals.

The introduction of artificial intelligence (AI) technologies into education dates back to the 1970s, when the idea arose of how researchers could provide an individual approach to learning using computers, which was the most effective and at the same time inaccessible to most people (Wilson & Scott, 2017: 2-19). The first attempts used artificial intelligence methods based on the rules of automatic adaptation or personalization of learning for each individual student (Bloom,1984:4-16).

Since then, the use of artificial intelligence technologies in education has developed in several directions. Starting with student-centered AI (i.e., tools to support learning and assessment), the end ends with teacher-centered AI (support for the learning process) and system-oriented AI (support for the management of educational institutions) (Carbonell, 1970: 190-202). In fact, the link between AI and education goes beyond using AI in classrooms (i.e. learning with AI), and also includes learning AI techniques (i.e. the study of AI technologies) and the preparation of citizens for life in the age of AI (i.e. cooperation with AI).

To fully understand the possibilities of AI technologies and reduce potential risks, as part of the educational policy, "how can we use artificial intelligence technology to improve education?" there is a need to create systematic answers to the main question.

The use of tools to support or improve AI-based learning has increased exponentially over the past decade (Zawacki-Richter et al., 2019:1-27). This process has become even more common due to the closure of schools during the Covid-19

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

pandemic. However, there is still insufficient data on how AI can help improve learning outcomes and help education theorists and practitioners better understand how effective learning occurs (Kreitmayer et al., 2018:1-4). To track learning outcomes in different situations, it is necessary to fully understand AI technologies, as well as evaluate competencies, especially technologies acquired in an informal and formal context.

Technologies demonstrate their potential in the process of monitoring the content of learning on various platforms based on the analysis of individual needs and the level of learning of students. An example of this is that one project aims to manage many thousands of open educational resources to make them accessible to all students (Berland et al., 2014: 1-10).

Materials and Methods

Higher education institutions have fewer data collection capabilities than online educational platforms. However, many universities have begun to implement automated Learning Management Systems (LMS) that allow data collection and storage. This shows the possibility of using machine learning technologies to improve the quality of education in higher education institutions.

The growth of data volumes contributed to the emergence in the early 2000s of a new direction in the field of artificial intelligence – the production of educational data (Educational Data Mining, EDM). EDM is a new field of research aimed at gaining knowledge from learning processes to support decision makers. The ability to collect data in LMS opens up new horizons for data mining researchers in the field of education, who can gain more in-depth information. Recently, EDM has been used within higher education to improve learning strategies. EDM involves the use of statistics, visualization, and machine learning techniques to explore and analyze educational data. EDM methods help in designing the educational environment, organizing course materials, and managing educational resources. These methods allow us to assess the impact of learning strategies implemented in certain educational institutions. In addition, EDM contributes to the development of the theory of learning developed by specialists in the field of educational psychology (West, 2012: 1-10).

EDM is closely related to learning analytics (LA). The goals and objectives of research in these disciplines largely coincide. LA is a new field of research aimed at improving the quality of education (Asif et al., 2017: 177-194). LA is an analytical approach aimed at analyzing, measuring and obtaining complete information about a student on various characteristics, including cognitive, social and psychological aspects, to help teachers make decisions about student achievements and failures (Burman & Som, 2019: 756-759).

Predictive analytics is an advanced field of analytics that predicts future results using statistical modeling, data mining techniques, and machine learning using integrated historical data. Predictive analytics is used for many learning purposes. In addition, work aimed at identifying students whose academic programs are at risk will

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

be an important part of predictive analytics. For example, the use of predictive models to show how the accuracy of forecasts changes as students move from course to course, and to determine whether they will finish the semester on time (Iqbal et al., 2019:1-7).

In the field of education, forecasting the size of tuition may be of interest (for example, student academic performance or skill acquisition), learning (for example, the impact of a given learning style or a particular teacher on a student), or other proximal values for organizations (for example, forecasts of retention or course registration). In education, predictive analytics is a well-established field of research, and currently some commercial products include predictive analytics in the educational content management system.

You should be able to distinguish predictive modeling from interpretative modeling. In interpretive modeling, the goal is to use all available data to interpret a given result. For example, in a regression model, observations of a student's age, gender, and socioeconomic status can be used to explain how this affects student academic performance. The purpose of these explanations is to test causal hypotheses (Shmueli, 2010: 289-310). The goal of predictive modeling activities is to create a model that predicts the values of new data (or a class if the forecast is not related to numerical data) based on observations. Predictive modeling differs from interpretive modeling in that it is based on the assumption that known datasets can be used to predict the value or class of new data based on observed variables.

Thus, the fundamental difference between interpretative modeling and predictive modeling is the application of the model to future events, which means that interpretative modeling does not seek to make any statements about the future, whereas predictive modeling does.

In the field of teaching and learning, predictive modeling covers a wide range of action-oriented educational policies and technological contexts in which educational institutions use these models in real time to meet the needs of students.

The purpose of predictive modeling is to set up a scenario that accurately characterizes the results of a given student. For example, a predictive model can be used to determine when a particular person will receive an academic degree. If an intervention strategy is not used, this model gives the individual student an idea of when they can get a degree in application.

Statistical modeling usually considers four types of data: categorical, ordinal, interval, and relative. The data differs in the types of relationships, that is, mathematical operations that can be obtained from individual elements. In practice, ordinal variables are often considered categorical, and the interval and ratio are considered numeric. Depending on the intended data types, there are two different classes of algorithms: classification algorithms for predicting categorical features and regression algorithms for predicting numerical features.

To build and apply a predictive model, it is necessary to select functions that correlate with the predictive value. There should be a single function that perfectly

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

correlates with the prediction of the selected result. However, this rarely happens in practice.

Some learning algorithms use all available attributes regardless of whether they have high information for prediction, while other models use some form of variable selection to exclude uninformative attributes.

Depending on the algorithm used to build the predictive model, it may be useful to examine correlations between signs or remove strong correlational elements, or conversion to functions is used to eliminate correlation.

By defining an information set of functions, it is possible to reduce the computational complexity of a predictive model, reduce data storage and collection requirements, and simplify the interpretation of predictive models.

A number of different well-known machine learning algorithms can be used to build predictive models with educational data. Them:

- With a controlled machine learning method: classification and regression, decision trees (DT), ensemble models(EM), K-nearest neighbors (KNN), Logistic regression (LR), Simple Bayesian (NB), random forest (RF) and auxiliary vector machines (SVM);
- With an unsupervised machine learning method: artificial neural networks (ANN), clustering (CLU) and correlation analysis algorithms.

Most EDM research focuses on classification and forecasting with large groups of students across many academic programs (Ojajuni et al.,2021: 481-491).

For a more detailed description of the algorithm for predicting students' academic performance, they go through the following stages:

1. Data collection and preparation. Data such as students' past achievements, information about the courses they are studying, and other factors that may affect their academic achievements should be collected. The data should be prepared for further use, that is, to clean up the data, remove missing values, bring the data into a single format, etc.
2. Model selection and comparison. At this stage, it is necessary to select the appropriate machine learning model that will be used to predict student performance. For example, it can be linear regression, a random forest, or a neural network. To achieve optimal accuracy, it is necessary to adjust the hyperparameters of the model.
3. Model training. At this stage, it is necessary to examine the model of the collected data. The data is broken down into training and test models, and the model is studied in the learning model. It is then tested in a test sample to evaluate the accuracy of the model.
4. Evaluation of the accuracy of the model. At the same time, the accuracy of the model is estimated based on indicators such as absolute average error (MAE) or quadratic average error (MSE).
5. Forecasting student academic performance. At this stage, the model is used to predict the student's academic performance in future courses. To do this, it is necessary

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

to enter data such as the student's past achievements and other factors that may affect his academic achievements.

6. Updating the model. If new data becomes available, the model can be updated to improve its accuracy and efficiency.

7. Interpretation of the results. At this stage, it is necessary to analyze the results obtained in predicting the student's academic performance in order to determine which factors affect the student's academic performance.

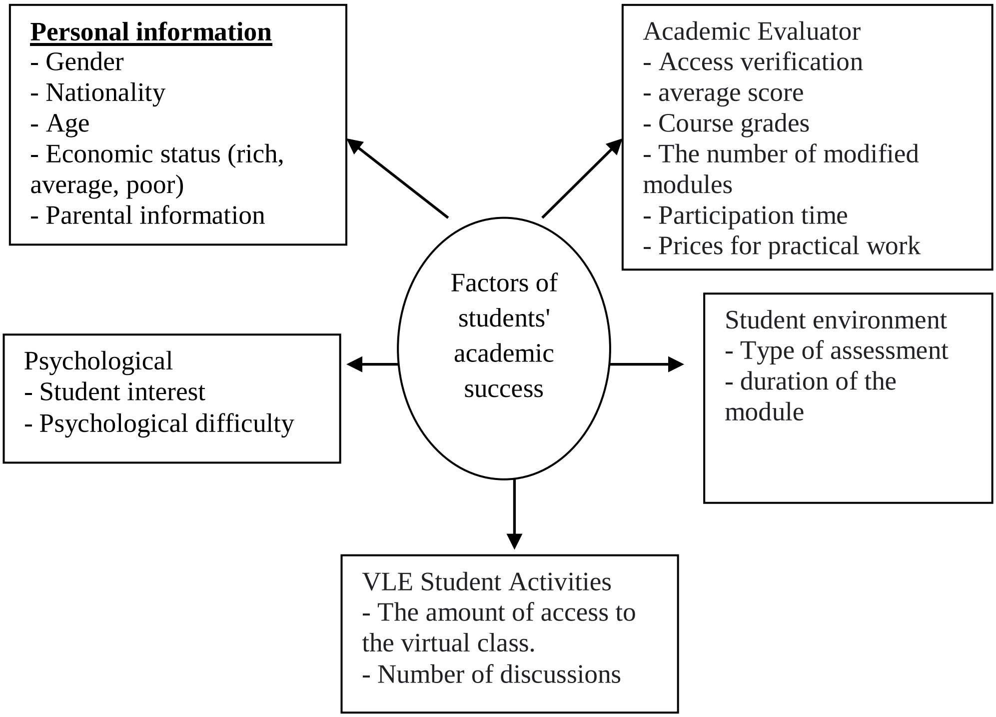


Figure 1. Attributes of predicting student success.

The algorithm for predicting progress includes several stages, from data collection and preparation to interpretation of the results. For successful forecasting, it is necessary to choose the right machine learning model, create and process the collected data (Fig.1).

Research results and analysis

To conduct the experiment, data were collected from 639 full-time students of the Department of Informatics of the O. Zhanibekov Sauth Kazakhstan Pedagogical University for the 2019-2023 academic year, the student grades were obtained, and class

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

attendance from University's learnig management system with the registrered name "Univer". (Fig. 2).

The purpose of the research experiment:

- Demonstration of the use of educational analytics (LA) with specific data sets;
- Predicting the performance of educational programs using machine learning algorithms.

One of the simplest and most useful machine learning models, linear regression, was used for the study. Linear regression is a machine learning algorithm used primarily for regression analysis. This model is a linear dependence function with one or more other variables (factors, regressors, independent variables).

A simple linear regression analysis finds a relationship between two continuous variables: the independent variable X (predictor) with the dependent variable y (resultant variable) on the axis. By constructing a regression line, one can generally predict the variable y from the equation of this line (URL, date of application: 13.02.2024).

$$y = w_0 + w_1 x \qquad (1)$$

Where x is the predictor variable, y is the value predicted by the corresponding value of x (the response variable). To determine the values of w0 and w1, the least squares formula is used to obtain the most suitable straight line. The calculations of w0 and w1 are shown as follows:

$$w_i = \frac{\sum_{i=1}^{|D|}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{|D|}(x_i - \overline{x})^2} w_0 = \overline{y} - w_1 \overline{x} \qquad (2)$$

After choosing a linear regression between machine learning methods that predict student grades, Python is loaded using Jupyter notebook, the program code below, using the NumPy programming language library, Pandas.

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

Жиынтық тізімдеме

факультеті: **Физика-математика** | Оқу сатысы: **бакалавриат** | Білім деңгейі : **Бірінші жоғары білім** | Курс: **3**
Оқу түрі: **күндізгі** | Мамандығы: **6B01503 - Информатика мұғалімін даярлау** | Бөлім: **Барлығы** | Семестр: **5**
оқу жылы: **2021-2022**

Ақы төлеу формасы: **Барлығы**
Жынысы: **Барлығы**

№	Аты-жөні	№ сынақ кітапшасы	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	к/д*	к/к*	Семестрге GPA		GPA курс		O G
1	Аба ***** Әді *****		0	56	80	85	52	61	92	58															8	30	1,96	C-	2,09	C	2,
2	Ади ***** Тал *****		0	84	100	90	76	82	78	82															8	30	2,96	B-	2,85	B-	2,
3	Ама ***** Ерл *****		0			75	62	80	60		72			60			71	70	71						10	39	2,64	C+	2,53	C+	2,
4	Амир ***** Керим *****		0	98	92	78	66	78	94	92															8	30	2,74	B-	2,77	B-	2,
5	Әбіш *****		0		92	65		70						94						94	90	90	76	70	10	39	2,98	B-	2,94	B-	2,
18	Мерге ***** али Диллаі *****		0	74	92	75	70	76	68	96															8	30	2,69	B-	2,71	B-	2,
19	Мирзаа ***** інур Турабас *****		0	97	98	86	84	86	68	96															8	30	2,93	B-	2,94	B-	3
20	Мы ***** ель Бау *****		0	88	86	85	82	77	78	96															8	30	2,71	B-	2,77	B-	2,
21	Нигма ***** нира Учкун *****		0	69	88	72	82	74	92	90															8	30	2,7	B-	2,67	B-	2,
22	Окт ***** нса Фар *****		0	68	92	80	94	80	84	82															8	30	2,78	B-	3,07	B	3,
23	Оспа ***** қызы		0	80	76	70	76	77	62	98					83	75									10	39	2,65	C+	2,76	B-	2,
24	Пай ***** одир Бобі *****		0	80	78	76	74	79	76	79															8	30	2,81	B-	2,71	B-	2,
25	Райь ***** а Сапа *****		0	93	96	76	89	73	76	88															8	30	2,7	B-	2,61	C+	2,
26	Рахм ***** Мей ***** айұлы		0	85	96	83	86	77	94	100															8	30	3,27	B	3,28	B	3,
27	Серик ***** Есимх *****		0	89	86	85	88	80	92	98															8	30	3,08	B	2,95	B-	3,
28	Умир ***** Ниша *****		0	96	90	77	60	78	94	58															8	30	2,12	C	2,13	C	2,
29	Урин ***** Мир *****		0	91	76	95	60	92	88	96															8	30	3,08	B	3	B	3,
30	Шай ***** кан Аянк *****		0	72	88	80	82	82	88	80															8	30	2,66	C+	2,52	C+	2,
31	Шар ***** зан Тем *****		0	80	71	85		79	94	66															7	28	2,73	B-	2,42	C+	2,
32	Шер ***** айё Пахр *****		0	75		80	72	75	56		68	60	76	80											10	39	2,4	C+	2,58	C+	2,
	Үздік		0	7	16	3	3	1	12	15	0	0	0	1	1	1	0	0	0	1	1	1	0	0							
	Жақсы		0	17	14	27	16	30	13	10	1	0	1	1	1	1	1	1	1	0	0	0	1	1							
	Сәтті		0	6	0	2	11	1	6	4	1	1	0	1	0	0	0	0	0	0	0	0	0	0							
	Қанағаттандырарлықсыз		32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0							
	Үлгерім		0	94	94	100	94	100	97	91	6	3	3	9	6	6	3	3	3	3	3	3	3	3							
	Сапалы санақ		0	75	94	94	59	97	78	78	3	0	3	6	6	6	3	3	3	3	3	3	3	3							

к/д* - Пәндер мөлшері
к/к* - Кредиттер саны

Студенттердің барлығы **32**
Олардың ішінен үлгерімі бар :
Өте жақсы: **0**

- [1, 1094]. - Дене шынықтыру
- [2, 544]. - Қазақстанның қазіргі заманғы тарихы
- [3, 1676]. - Қазақ (орыс) тілі
- [4, 548]. - Математика 1
- [5, 16]. - Саясаттану
- [6, 34]. - Физика

Жақсы: **29**
Орташа: **3**
Қанағаттанарлықсыз: **0**

- [7, 25]. - Әлеуметтану
- [8, 7205]. - Шетел тілі
- [9, 956]. - Қазақ (орыс) тілі
- [10, 7046]. - Сандар теориясы
- [11, 93]. - Элементарлық математика
- [12, 3483]. - Шетел тілі
- [13, 3769]. - Элементарлық математика
- [14, 7207]. - Алгебра және сандар теориясы

Figure 2. Initial data on student academic performance.

From the LMS database, the results of students' education according to the educational program for 6B01503 - computer science teacher training was selected, that

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

is, their academic performance during the semester, the effectiveness of the session. The data can be downloaded in MS Excel format (Fig.3).

For data analysis and transformation, the data is cleared based on the selected data. The results of study in subjects are selected as a predicate, i.e. information about general education disciplines, basic disciplines, specialized disciplines and the average bachelor's degree score (GPA) is summarized and performed in an Excel spreadsheet. This data was saved in CSV format for insertion into machine learning methods (Fig.3).

At the prediction stage, datasets are loaded from a CSV file, splits the data into groups for reading and testing, creates a machine learning model using linear regression, compares the model with the training data and uses a model that predicts the target variable for the test data, the main parameters of which are calculated student performance data.

The objective linear regression function consists of one feature X, where academic performance scores are one of the cycles of general education disciplines, basic disciplines, specialized disciplines, and the target variable is the average bachelor's degree score (GPA) (Fig.4).

Full name	Political science	Cultural studies	Social recognition	Modern history of Kazakhstan	Programming language Python	GPA
Ab*****naral	96	99	99	99	85	3,18
I*****a	97	100	100	100	97	3,51
Ba*********kyzy	99	98	97	100	78	3,05
Bal*************na	98	100	99	99	88	3,17
Ba*********nuly	96	99	99	100	72	2,82
Be*********ovna	100	99	100	99	95	3,12
Ge*********vich	84	70	87	58	84	2,96
Bim*********itkakyzy	100	100	98	99	75	3,37
Dos************a	98	99	100	98	86	3,13
Yem************uly	100	99	95	100	87	2,9
aqs*********kq?z?	100	100	99	99	85	3,07
an*********atkizy	100	99	97	99	100	3,13
Zhol*********liyarkyzy	100	99	98	99	91	3,11
Zhu*********nurkyzy	95	100	98	98	95	3,05
Zhu*********nkyzy	90	83	85	100	84	3,28
Ka*********uly	98	100	97	98	76	2,73
K*********a	98	80	90	95	95	3,08
Kur************uly	99	100	75	98	95	2,86
Ma*********tovna	94	99	91	97	88	3,01
Nam*********satkyzy	93	100	96	100	82	3
Per*********ovich	100	100	98	100	81	2,94
Pola*********aruly	100	98	94	98	65	2,84
i*********ch	97	100	97	98	75	2,84
Turl*********ayevna	99	99	98	100	83	3,44
Us*********zy	92	97	98	98	78	3,14

Figure 3. average student score in CSV file.

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

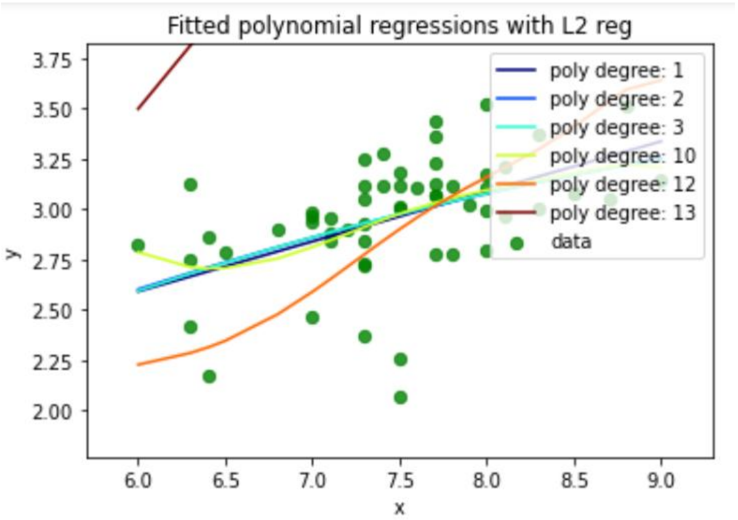
Ввод [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
data = pd.read_csv('RegressionL8.csv', sep=';')
data.head()
```

Out[2]:

	X1	Y
0	52	2,19
1	56	2,58
2	56	2,09
3	58	2,96
4	58	2,33

Figure 4. Features and target variable of the linear regression objective function.

Python was used to demonstrate experiments and obtain graphs (Fig.5). Additional constraints were added to the problem conditions in order to reduce the error and the number of variables used in order to create an algorithm to eliminate overfitting. In the study, L1 regression (Lasso regression) and L2 ridge regression (ridge regression) (URL, date of application: 13.02.2024), (Fig.6).



ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

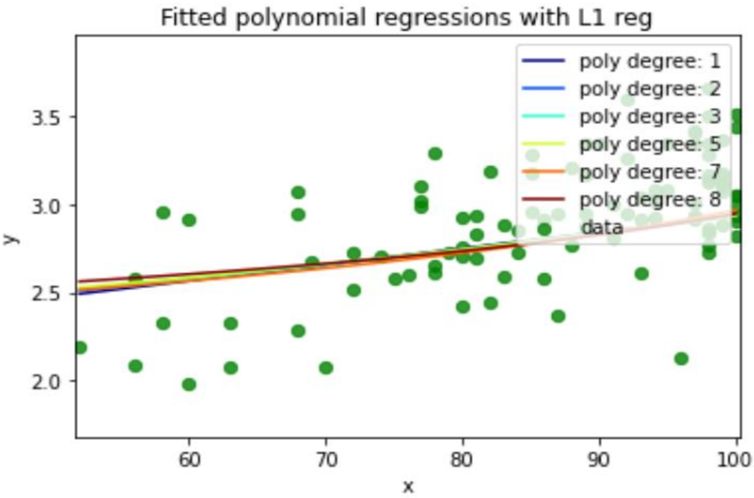
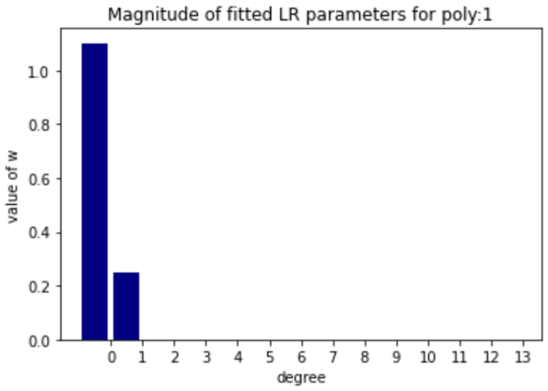
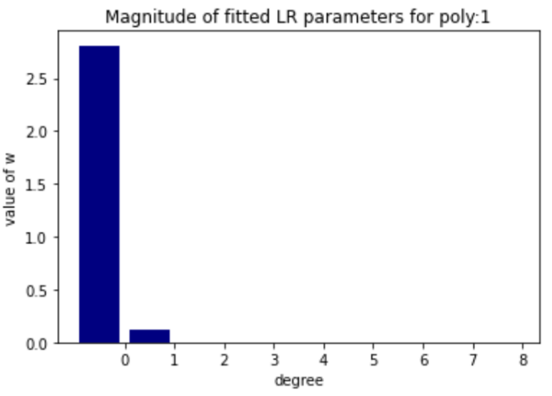


Figure 5. Polynomial regressions with L2 and L1 regularization.

The free coefficients of the empirical linear regression model with L1 and L2 regularization are presented in the form of a graph.



a)



б)

Figure 6. a) free coefficients of the objective function of linear regression with regularization L2 (in ridge regression) w_0 = 1.1, w_1 = 0.25 result of training in a cycle of basic disciplines.

b) L1 in the cycle of general education disciplines (in Lasso regression) w_0 = 2.8, w_1 = 0.12 the result of training in the cycle of general education disciplines.

The linear regression model helps to predict the value of the dependent variable, and also helps to explain how accurate the forecast is. This is determined by the values of the parameters of the free coefficients. The value of w_1 indicates which part of the variation of the dependent variable can be explained by the independent variable. The values of w_1 range from 0 to 1. A value of 0.12 means that the independent variable can explain 12 percent of the change in the observed values of the dependent variable. A value of 1 means that in practice it is possible to make an excellent assumption, which is rarely found. A value of 0 means that the independent variable does not help to predict the dependent variable at all.

A linear regression model of the forecast $y=2.8+0.12x$ was obtained and it was found that the results of teaching "modern history of Kazakhstan" from the cycle of general education disciplines affect student academic performance by 12%. In addition, it can be concluded that the results of teaching basic subjects such as "Math1", "Math2" affect the results of students' GPA by 25% according to the prediction model with $y=1.31+0.25x$. The results of teaching specialized disciplines "programming languages", "methods of teaching computer science" affect the academic performance of students according to the model $y=1.2+0.32x$ by 32% (Fig.7).

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

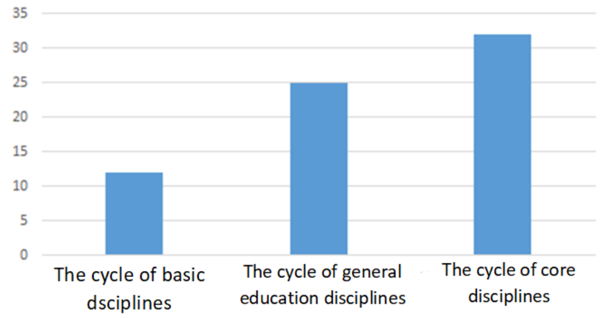


Figure 7. The result of the experiment.

The results of this study can be used in the development of an educational program and forecasting student academic performance. In summary, it helps integrate learning analytics into LMS and reinforces the importance of developing educational programs. It also helps to understand how feature selection can improve the performance of machine learning models.

Conclusion

Learning Analytics (LA) and educational Data mining (EDM) are receiving increasing attention from academic communities by providing educational data interventions. In addition, the advantages of educational analytics and its integration into LMS are now widely known. This allowed the University to take the necessary preventive measures, providing early notification in real time regarding the participation and academic performance of students. In this article, machine learning methods were used to integrate learning analytics into LMS and demonstrate the process of predicting the implementation of an educational program. Understanding linear regression is the key to understanding complex models down to deep neural networks. Additional features such as student usage data, attendance records, etc. are combined for other assumptions about the quality of future work.

References

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & education, 113, 177-194. https://doi.org/10.1016/j.compedu.2017.05.007

Burman, I., & Som, S. (2019, February). Predicting students academic performance using support vector machine. In 2019 Amity international conference on artificial intelligence (AICAI) (pp. 756-759). IEEE. DOI:10.1109/AICAI.2019.8701260

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. Technology, Knowledge and Learning, 19, 205-220.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational researcher, 13(6), 4-16. https://doi.org/10.3102/0013189X01300600

Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. IEEE transactions on man-machine systems, 11(4), 190-202. DOI:10.1109/TMMS.1970.299942

Iqbal, Z., Qayyum, A., Latif, S., & Qadir, J. (2019, February). Early student grade prediction: an empirical study. In 2019 2nd international conference on advancements in computational sciences (ICACS) (pp. 1-7). IEEE. DOI:10.23919/ICACS.2019.8689136

Kreitmayer, S., Rogers, Y., Yilmaz, E., & Shawe-Taylor, J. (2018). Design in the Wild: Interfacing the OER learning journey. ACM: Association for Computing Machinery. DOI:10.14236/ewic/HCI2018.165

Ojajuni, O., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., ... & Mbarika, V. (2021). Predicting student academic performance using machine learning. In Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13-16, 2021, Proceedings, Part IX 21 (pp. 481-491). Springer International Publishing.

Rossman, A. J., & Chance, B. L. (2014). Using simulation-based inference for learning introductory statistics. Wiley Interdisciplinary Reviews: Computational Statistics, 6(4), 211-221. https://doi.org/10.1002/wics.1302

Shmueli, G. (2010). To explain or to predict? Publisher: JSTOR, pp. 289-310.

URL: Basic principles of machine learning on the example of linear regression. Available: https://habr.com/ru/companies/ods/articles/322076/ (date of application: 13.02.2024).

Wilson, C., & Scott, B. (2017). Adaptive systems in education: a review and conceptual unification. The International Journal of Information and Learning Technology, 34(1), 2-19. https://doi.org/10.1108/IJILT-09-2016-0040

West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. Governance studies at Brookings, 4(1), 1-10.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International Journal of Educational Technology in Higher Education, 16(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Нахипова Венера Исмаиловна[1], Сулейменова Лаура Аскарбековна[2], Адылбекова Эльвира Тулепбергеновна[3].
[1]Докторант, [2]Кандидат технических наук, [3]Кандидат педагогических наук.
[1,2,3] Южно-Казахстанский педагогический университет имени О. Жанибекова. г. Шымкент, Республика Казахстан.

Определение успеваемости обучающихся с использованием методов машинного обучения

Аннотация. В исследовательской работе рассматриваются важные аспекты использования подходов к аналитике обучения (LA) и анализу образовательных данных (EDM), которые можно использовать для разработки образовательных программ и отслеживания прогресса. Было показано, что интеграция аналитики обучения в систему управления обучением (LMS) может повысить эффективность образовательного процесса за счет прогнозирования и предотвращения проблем, с которыми сталкиваются студенты в отношении успеваемости. Рассматривается процесс анализа и прогнозирования данных по успеваемости с использованием методов машинного обучения, в том числе линейной регрессии, с определением соотношения успеваемости по дисциплине и баллов GPA. Результат исследования показывает важность выбора правильных функций для повышения точности прогнозирования с помощью моделей машинного обучения, а также обеспечивает предоставление входных данных для улучшения образовательной программы. Рассматривается процесс создания прогнозных моделей для оценки успеваемости учащихся путем сбора и подготовки данных с использованием алгоритмов машинного обучения. Дается четкое описание использования моделей для прогнозирования будущей успеваемости учащихся, а также особое внимание уделяется интерпретации полученных результатов для выявления основных факторов, влияющих на успеваемость. Эта работа способствует развитию области аналитики обучения и областей производства образовательных данных, предлагая практические рекомендации по управлению учебными программами и повышению качества образования на основе данных.

Ключевые слова: машинное обучение, искусственный интеллект, LMS, аналитика обучения, интеллектуальный анализ образовательных данных, предиктивная аналитика, линейная регрессия**.**

ISSN 3078-8285 (Print), ISSN 3078-8293 (Online). https://ilim.okmpu.kz/

Нахипова Венера Исмаилқызы[1], Сүлейменова Лаура Аскарбекқызы[2], Адылбекова Эльвира Тулепбергеновна[3].

[1]Докторант, [2]Техника ғылымдарының кандидаты, [3]Педагогика ғылымдарының кандидаты.

[1, 2, 3] Ө.Жәнібеков атындағы Оңтүстік Қазақстан педагогикалық университеті, Шымкент қ., Қазақстан.

Машиналық оқыту әдістерін қолдану арқылы білім алушылардың үлгерімін анықтау

Аннотация. Зерттеу жұмысында білім беру бағдарламаларын әзірлеу мен үлгерімді бақылау үшін қолдануға болатын оқу аналитикасы (LA) мен білім беру деректерін талдау (EDM) жолдарын пайдаланудың маңызды аспектілері қарастырылады. Оқу аналитикасын оқытуды басқару жүйесіне (LMS) біріктіру арқылы студенттердің оқу үлгеріміне қатысты кездесетін қиындықтарды болжау және алдын алу арқылы білім беру үдерісінің тиімділігін арттыруға болатындығы келтірілген. Машиналық оқыту әдістеріне, оның ішінде сызықтық регрессияны қолданып пән бойынша оқу үлгерімі мен GPA балл байланысын анықтап, оқу үлгерімі бойынша деректерді талдау және болжау үдерісі қарастырылады. Зерттеу нәтижесі машиналық оқыту модельдерімен болжаудың дәлдәгән жоғарылату үшін дұрыс функцияларды таңдау маңыздылыңын көрсетеді, сонымен қатар білім беру бағдарламасын жақсатру үшін кіріс берілгендерін қамтамасыз етеді. Машиналық оқыту алгоритмдерін қолдана отырып, деректерді жинау мен дайындау арқылы студенттердің оқу үлгерімін бағалау үшін болжамды модельдер құру үдерісі қарастырылады. Студенттердің болашақтағы үлгерімін болжау үшін модельдерді қолданудың нақты сипаттамасы келтірілген, сондай-ақ оқу үлгеріміне әсер ететін негізгі факторларды анықтау үшін алынған нәтижелерді түсіндіруге ерекше назар аударылады. Бұл жұмыс берілгендер негізінде оқу бағдарламаларын басқару мен білім беру сапасын көтеру үшін практикалық нұсқаулар ұсына отырып, оқу аналитикасы саласы мен білім беру деректерін өндіру салаларын дамытуға үлес қосады.

Кілт сөздер: машиналық оқыту, жасанды интеллект, LMS, оқу аналитикасы, білім беру деректерін өндіру, болжамдық аналитика, сызықтық регрессия.