

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ /  
MATHEMATICAL MODELING, NUMERICAL METHODS AND PROGRAM COMPLEXES

DOI: <https://doi.org/10.23670/IRJ.2023.137.15>

МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ЭФФЕКТИВНОСТИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА В  
УНИВЕРСИТЕТЕ

Научная статья

Дорофеева Ю.А.<sup>1,\*</sup>, Правдин К.В.<sup>2</sup>, Михайлов В.И.<sup>3</sup>

<sup>1</sup>ORCID : 0000-0002-9327-6705;

<sup>1,2,3</sup>Национальный исследовательский университет ИТМО, Санкт-Петербург, Российская Федерация

\* Корреспондирующий автор (julana2008[at]yandex.ru)

**Аннотация**

Система высшего образования динамично развивается с каждым годом, что приводит к естественному конкурированию вузов друг с другом. Одной из компонент оценивания деятельности университета является оценка качества предоставляемого образования. Высокие показатели, попадание в авторитетные рейтинги напрямую зависят от количества выпускников и качества их профессионализма.

В работе будет рассмотрена математическая модель повышения эффективности образовательного процесса при реализации факультативных дисциплин на примере реальных данных. В ходе исследования определена вероятность окончания вуза студентами, а также проведён сравнительный анализ применения моделей машинного обучения в процессе прогнозирования.

Также в статье представлены результаты численного моделирования, полученные при использовании языка программирования Python и дополнительных библиотек анализа данных.

**Ключевые слова:** оптимизация, моделирование, интеллектуальный анализ данных, машинное обучение в образовании, академическая успеваемость, система управления обучением, предсказания, линейное программирование.

MODELLING AND FORECASTING THE EFFECTIVENESS OF THE EDUCATIONAL PROCESS AT  
UNIVERSITIES

Research article

Dorofeeva Y.A.<sup>1,\*</sup>, Pravdin K.V.<sup>2</sup>, Mikhailov V.I.<sup>3</sup>

<sup>1</sup>ORCID : 0000-0002-9327-6705;

<sup>1,2,3</sup>ITMO National Research University, Saint-Petersburg, Russian Federation

\* Corresponding author (julana2008[at]yandex.ru)

**Abstract**

The system of higher education is dynamically developing every year, which leads to natural competition of universities with each other. One of the components of university activity evaluation is the assessment of the quality of education provided. High indicators, getting into authoritative ratings, directly depend on the number of graduates and the quality of their professionalism.

The work will review the mathematical model of improving the efficiency of the educational process in the implementation of elective disciplines on the example of real data. In the course of the study, the probability of students graduating from university is determined, and a comparative analysis of the application of machine learning models in the forecasting process is carried out.

The article also presents the results of numerical simulations obtained using the Python programming language and additional data analysis libraries.

**Keywords:** optimization, modelling, data mining, machine learning in education, academic performance, learning management system, predictions, linear programming.

**Введение**

Относительно большое количество студентов не получают диплом об образовании в процессе обучения в вузе [1]. Это касается не только студентов Российской Федерации, но и стран Европы и США. Причин этому явлению несколько. Самой основной и распространённой является академическая неуспеваемость студентов, которая может возникнуть не только по причине отсутствия знаний по какому-либо предмету, но и из-за систематических нарушений учебной дисциплины. Ещё одним фактором являются жизненные обстоятельства, которые включают в себя изменение места жительства, социального статуса, семейного положения и т. д.

Система высшего образования в целом ориентирована на обеспечение высокого образовательного стандарта. Для достижения этой цели необходим строгий административный подход. Разработка образовательных программ, набор абитуриентов, кадровое и материально-техническое обеспечение, финансы и т. д. Однако именно успешная обучаемость студентов является основным показателем эффективности работы высших учебных заведений. Именно она служит главным критерием для оценки качества учебного процесса. В связи с этим одной из важных проблем является анализ причин, влияющих на отчисление, а также прогнозирование успешного окончания вуза [2].

С целью расширения научного кругозора, улучшения качества прикладных знаний студентов, во многих вузах реализуются дополнительные факультативные дисциплины по выбору. Данная форма занятий является относительно гибкой как с точки зрения методики, так с точки зрения и содержания. Это может быть классическое занятие, например, лекция или семинар, а может и свободная дискуссия. Выборность реализуется в начале каждого семестра и осуществляется по установленным университетом правилам. Однако существует такая проблема, как неоптимальное использование выделенных бюджетных средств вуза на реализацию спецкурсов, так как определенное количество студентов, изъявивших желание посещать курс, так его и не оканчивают. Причины этого могут быть различны. Важно отметить, что в предложенной модели факультатив представляет собой особый вид занятий (с применением различных педагогических технологий).

На данный момент существует достаточное количество моделей, описывающих те или иные процессы обучения, эффективности, оптимизации и т.д. С изменением образовательной среды менялись и актуальные методы автоматизированного анализа данных в образовательных системах. Так, в работах [3], [4], [6], [7] представлены математические модели и методы моделирования показателей эффективности деятельности вуза. В научных трудах [8], [9] рассматривались сущности прогнозирования результатов обучения, основанные на корреляционном, регрессионном, кластерном и факторном анализах. Кроме этого, с помощью автоматизированного системно-когнитивного анализа и его программного инструментария – интеллектуальной системы «Эйдос» в монографии [10] была изучена зависимость учебных достижений студентов университета от их занятий в учебном подразделении в период обучения в средней школе и от результатов ЕГЭ. Данный метод является достаточно обоснованным и хорошо изученным в основе которого эффективная модель, относящаяся к нечетким декларативным гибридным моделям.

В новых исследованиях широко используются методы EPM (Education Process Mining) и машинного обучения (Machine Learning – ML) с целью анализа образовательных данных [11].

### Постановка задачи

В связи с вышеизложенным была создана математическая модель прогнозирования эффективности образовательного процесса в вузе с учетом применения дополнительных образовательных технологий.

Для достижения поставленной цели необходимо решить следующие задачи:

- создание оптимизационной модели на основе машинного обучения для прогнозирования вероятностей окончания университета по заранее выбранным метрикам качества;
- сравнительный анализ уравнений регрессии (логистическая, гребневая, Лассо);
- численное моделирование алгоритма с оптимальным уравнением регрессии при использовании языка Python;
- вывод об оптимальности использования регрессионных соотношений с точки зрения точности прогноза.

### Основные результаты

Пусть в университете обучается  $N$  студентов. Обозначим за  $N_q$  множество тех студентов, кто принял решение записаться на дополнительные дисциплины:  $N_q \subset N$ . Количество записавшихся на дополнительные дисциплины студентов будет равно мощности множества  $N_q$ , то есть  $n = |N_q|$

Введём понятие рейтинга вуза, под которым будем понимать оценку деятельности образовательного учреждения на основании различных учебных активностей и мероприятий. Понятно, что общий (интегральный) рейтинг – линейная комбинация частных ( $L_i$ ), взятых с априори заданными весовыми коэффициентами.

В данной работе рассмотрим частный рейтинг вуза на основании организации факультативных дисциплин. Пусть он будет вычисляться как произведение количества студентов  $x_j$  и вероятностей их выпуска  $p_j$ :

$$L_{\text{фак}}(x) = \sum_{j=1}^n p_j x_j \rightarrow \max$$

На реализацию одной дополнительной дисциплины университет выделяет  $b_i$  рублей. Под  $a_{ij}$  понимается количество затраченных средств на использование образовательных технологий для  $j$ -го студента (в предложенной постановке это факультативные занятия). Тогда основной целью для вуза будет повышение вероятности выпуска студентов при минимизации затрат на обеспечение. Постановка задачи принимает вид линейного программирования при ограничениях:

$$\begin{cases} \sum_{i=1}^n a_{ij} x_j \leq b_i & i = \overline{1, \dots, m}, j = \overline{1, \dots, n} \\ x_j \geq 0 \end{cases}$$

Примем во внимание тот факт, что число мест на каждом курсе ограничено, то есть мощность множества студентов на факультативе не превосходит наперед заданного числа:  $|N_{qi}| \leq c, i = \overline{1, \dots, m}$ .

В данной постановке существует два сценария:

- любой студент может записаться на дополнительный понравившийся курс;
- выбор курса предоставляется лицам с вероятностью окончания университета не менее 0,5.

Для проверки эффективности и адекватности модели были проведены вычисления, в ходе которых была рассмотрена группа из 100 студентов, охватывающая 63 различных факультатива. Количество участников на каждом курсе не превышало 30. Стоимости курсов  $b_i$  варьировались в пределах 2000-2500 евро, стоимость обеспечения техникой  $a_{ij}$  в пределах 15-150 евро соответственно.

Получены значения функционала равны

$$L_{\text{фак}}^1 \approx 63,94$$

при произвольной вероятности  $p_j$  окончания университета и

$$L_{\text{фак}}^2 \approx 89,47$$

при вероятности участия на курсе только тех студентов, у кого  $p_j \geq 0,5$ .

На основании полученных результатов можно сделать вывод о целесообразности записи на факультативы только студентов с относительно высоким уровнем обученности.

Далее в исследовании прогнозируется вероятность окончания высшего учебного заведения на реальном открытом наборе данных, предоставленным отделом академических услуг Политехнического университета Порталегре [12]. Здесь учитывается фактор дополнительных образовательных технологий, в том числе и в форме дополнительных занятий со студентами с применением различных методик.

### Описание исследуемых данных

Набор данных включает в себя демографические, социально-экономические и макроэкономические сведения, известные на момент зачисления студентов, а также текущие результаты окончания первого и второго семестров, собранные за период с 2008 по 2019 учебный год. Данные представлены в виде табличного текстового файла CSV, содержащего 3630 строк по 35 столбцов в каждой.

В выборке известны апостериорные вероятности статуса студента, принимающие два возможных значения  $y_i \in \{0, 1\}$ . Анализируя исходные данные установлено, что 1421 учащийся были отчислены, а 2209 продолжили своё обучение (см. рисунок 1).

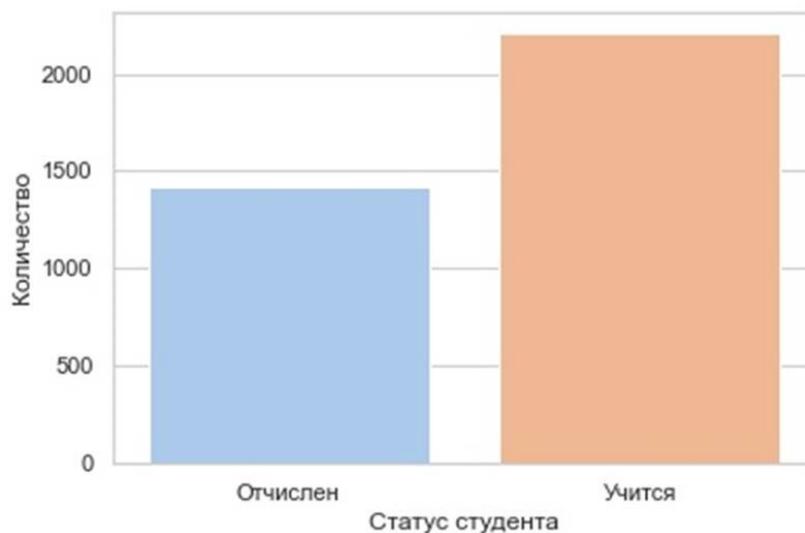


Рисунок 1 - График распределения студентов  
DOI: <https://doi.org/10.23670/IRJ.2023.137.15.1>

### Отбор факторов (образовательных технологий) и методы построения множественной регрессии

Для успешного прогнозирования вероятностей необходимо установить функциональную связь между зависимыми переменными  $y = (y_1, y_2, \dots, y_{34})^T$  – вероятностями окончить университет, для которых применяется соответствующая образовательная технология, и группой независимых переменных  $x = (x_1, x_2, \dots, x_{34})^T$  – количества студентов, для которых применяется соответствующая образовательная технология, с учётом ошибки модели  $e$  – случайной величины, характеризующей отклонения результативного признака от теоретического, найденного по уравнению регрессии:

$$p_j = f_j(x_1, x_2, \dots, x_{34}) + e$$

В данном случае рассматриваются три часто используемые модели машинного обучения: логистическую, гребневую и Лассо регрессию. В дальнейшем необходимо провести сравнительный анализ их применения к выбранным данным.

1) В логистической регрессии рассматривается модель следующего вида:

$$\hat{f}(x) = \sigma \left( \sum_{i=0}^{34} \omega_i x_i \right)$$

где  $\sigma(x) = \frac{1}{1+e^{-x}}$  – сигмоидная функция. Для определения коэффициентов регрессии используется метод максимального правдоподобия. Тогда функция минимизации принимает следующий вид:

$$\mathcal{L}(w) = -\frac{1}{n} \sum_{i=1}^n \ln P(y = y_i | x_i) \rightarrow \min_w$$

$$\ln P(y = y_i | x_i) = y_i \ln \left( \hat{f}(x_i) \right) + (1 - y_i) \ln \left( 1 - \hat{f}(x_i) \right)$$

В качестве метода поиска экстремума применяется градиентный спуск, в котором пересчитываются веса по формуле  $w_{i+1} := w_i - \alpha \frac{\partial \mathcal{L}}{\partial w}$ , где  $\alpha$  – шаг градиента.

2) В *гребневой регрессии* рассматривается линейная зависимость  $f(x, \omega) = \langle \omega, x \rangle$ . При этом накладываются ограничения на величину коэффициентов  $\omega : \|\omega\|_2^2 \leq t^2$ . Минимизирующий функционал  $\mathcal{L}$  с учётом ограничений принимает вид:

$$\mathcal{L}_\lambda(\omega) = \|F\omega - y\|^2 + \lambda\|\omega\|^2$$

где  $\lambda$  – неотрицательный параметр. Решением в этом случае будет вектор  $\omega^* = (F^T F + \lambda I_n)^{-1} F^T y$ , где  $\lambda I_n$  – диагональная матрица, называемая гребнем.

3) В *регрессии Лассо* используется другое ограничение на коэффициенты  $\omega : \|\omega\|_1 \leq t$ . Положительным моментом использования  $l_1$  нормы является не только осуществление регуляризации, но и приравнение некоторых коэффициентов  $\omega$  к нулю при достаточно большом значении  $\lambda$ . Минимизирующий функционал такой же, как и в гребневой регрессии.

Для сравнения и оценки качества применимы следующие метрики:

– среднеквадратическая ошибка

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

– средняя абсолютная ошибка

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

– коэффициент детерминации  $R^2$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_l - y_i)^2}$$

Для реализации численного моделирования использовался язык программирования Python в совокупности с библиотекой машинного обучения Scikit-learn. Выборка данных была предварительно разделена на обучающую и тестовую в соотношении 7:3. В результате проведения экспериментов были получены результаты, представленные в таблице 1. Наилучшие показатели в рамках текущей постановки задачи продемонстрировала логистическая регрессия.

Кривая обучения модели в зависимости от объёма обучающего набора данных представлена на рисунке 2.

Таблица 1 - Результаты численного моделирования множественной регрессии

DOI: <https://doi.org/10.23670/IRJ.2023.137.15.2>

Оценка	Модель		
	Логистическая регрессия	Гребневая регрессия	Лассо регрессия
RMSE	0,2486	0,2777	0,2957
MAE	0,1301	0,2002	0,2179
$R^2$	0,6407	0,6776	0,6344

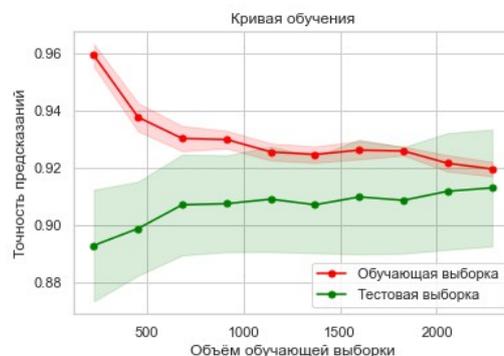


Рисунок 2 - График обучения модели логистической регрессии

DOI: <https://doi.org/10.23670/IRJ.2023.137.15.3>

В ходе обучения сформировался следующий вектор весовых коэффициентов:

$$w = [0,2974 - 0,0336 - 0,0278 - 0,1090 - 0,3350 \ 0,0197 \\ 0,0127 - 0,0118 \ 0,0097 \ 0,0526 - 0,0036 - 0,2874 \\ - 0,0351 - 1,3053 \ 2,1876 - 0,3922 \ 0,8138 - 0,0362]$$

0,1512 – 0,1962 – 0,2975 – 0,0040 0,6590– 0,1134  
 0,0969 – 0,2003 – 0,6431 – 0,0303 0,9486 0,1063  
 0,2569 – 0,0763 0,0101 – 0,0087]

Существенность полученных коэффициентов определяется отношением  $t_i = w_i/\sigma_i, i = \overline{1, \dots, 34}$ , где  $w_i$  – коэффициент множественной регрессии;  $\sigma_i$  – среднее квадратическое отклонение этого коэффициента. Если  $t_i < t_{\text{табл}}$ , взятого по таблицам  $t$ -распределения Стьюдента, то с заданной вероятностью не отвергается гипотеза, что соответствующий коэффициент регрессии в генеральной совокупности приравнивается к нулю. При этом  $i$ -й фактор в таком случае признаётся несущественным для построения регрессионного уравнения. Если при проведении исследования оказывается, что вычисленные значения  $t_i$  для нескольких факторов не превышают  $t_{\text{табл}}$ , то факторы исключаются поочередно, начиная с наименьшего по абсолютной величине  $t_i$ . Процесс исключения происходит до тех пор, пока не будет выполняться соотношение  $t_{\text{min}} \geq t_{\text{табл}}$ .

При уровне значимости  $\varepsilon = 0,05$  было установлено, что вектор весовых коэффициентов можно сократить до 9 факторов, вместо предполагаемых 34 с незначительной потерей качества  $e = 10^{-3}$ , с заметным упрощением модели, а также уменьшением мультиколлинеарности в наборе данных между сильно коррелированными признаками.

### Практический аспект

Модель прогнозирования эффективности образовательного процесса может быть использована для определения того, какие методы и подходы в образовании наиболее эффективны в процессе достижения желаемых целей. Например, она может помочь школам и университетам определить, какие курсы и программы наиболее успешны, а также какие учебные материалы и методы преподавания наиболее эффективны.

Представленная модель может быть использована для прогнозирования будущих результатов обучения студентов, на основе их текущих оценок. В этом случае преподавателям и администрации образовательных учреждений можно будет определить, какие студенты нуждаются в дополнительной поддержке и помощи, чтобы достичь успеха в учебе.

Кроме того, появляется возможность анализа данных об успеваемости и приоритетных направлениях для студентов (на основе выборности дополнительных дисциплин). Это дает возможность образовательным учреждениям разрабатывать более эффективные стратегии обучения и улучшать качество образования.

В целом модель прогнозирования эффективности образовательного процесса является востребованным инструментом для улучшения качества образования и достижения лучших результатов в учебе.

### Заключение

В ходе качественного анализа сравнения моделей машинного обучения для отсеивания студентов был реализован алгоритм логистической регрессии, показавший хорошие результаты на практике. Предложенный подход универсален и может использоваться в любом учреждении высшего или специального профессионального образования. Применение модели поможет выявить основные причины отчисления обучающихся и скорректировать работу вуза в дальнейшем. Разработанная модель послужила методической основой для реализации общеуниверситетских факультативных дисциплин, где возможность записи на курс будет вычисляться на основе полученных вероятностей. Результаты данного исследования имеют практическую значимость для лиц, принимающих управленческие решения о формировании стратегии развития и конкурентоспособности высших учебных заведений в условиях ограниченного бюджета.

Тем не менее выбранный подход не является совершенным. В дальнейшем планируется улучшение используемой модели с учетом методики выбора дополнительных занятий, а также применением дополнительных технологий. Модель не рассчитана на прогноз относительно основных дисциплин, что также будет учитываться в дальнейшей работе над ней.

Программный код опубликован в Интернет-репозитории [13].

### Конфликт интересов

Не указан.

### Conflict of Interest

None declared.

### Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

### Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

### Список литературы / References

1. Официальный сайт Национального университета «Высшая школа экономики». — URL: <https://www.hse.ru/> (дата обращения: 12.09.2023)
2. Realinho V. Predicting Student Dropout and Academic Success / V. Realinho, J. Machado, L. Bartista et al. // Data. — 2020. — V. 7. — P. 146. — DOI: 10.3390/data7110146
3. Зятева О.А. Интеллектуальный анализ данных при категоризации преподавателей вуза на основе наукометрических показателей / О.А. Зятева, Е.А. Питухин, И.В. Пешкова и др. // Инженерный вестник Дона. — 2017. — №4. — URL: <https://ivdon.ru/magazine/archive/n4y2017/4580> (дата обращения: 12.09.2023)

4. Котенко Ю.С. Проблемы современного вуза и маркетинговые методы их выявления и оценки / Ю.С. Котенко, И.А. Названова, М.Г. Подопригора и др. // Инженерный вестник Дона. — 2013. — №2. — URL: <https://ivdon.ru/ru/magazine/archive/n2y2013/1631> (дата обращения: 12.09.2023)
5. Яндыбаева Н.В. Разработка программного продукта для определения эффективности деятельности высшего учебного заведения / Н.В. Яндыбаева, Е.Р. Кожанова, В.А. Кушников // Вестник Саратовского государственного технического университета. — 2014. — Т. 75, №2. — С. 214-219. — URL: <https://elibrary.ru/item.asp?id=22567751> (дата обращения: 12.09.2023)
6. Яндыбаева Н.В. Математические модели, алгоритмы и комплексы программ для мониторинга эффективности образовательной деятельности вуза / Н.В. Яндыбаева, В.А. Кушников // Проблемы управления. — 2015. — №1. — С. 53–63. — URL: <https://elibrary.ru/item.asp?id=22958726> (дата обращения: 12.09.2023)
7. Яндыбаева Н.В. Моделирование и прогнозирование показателей эффективности образовательной деятельности высшего учебного заведения / Н.В. Яндыбаева // Вестник Мордовского университета. — 2018. — Т. 28, № 1. — С. 120–136. — DOI: 10.15507/0236-2910.028.201801.120-136.
8. Майер Р.В. Классификация учебных факторов методом кластерного анализа / Р.В. Майер // Проблемы учебного физического эксперимента: сб. науч. и метод. работ. — 1998. — Вып. 5. — С.12-19.
9. Шевченко В.А. Концепция построения модели приобретения знаний студентами по дисциплине «Информатика» / В.А. Шевченко // Вестник ХНАДУ: сб. науч. тр. — 2012. — Вып. 56. — С. 26–29.
10. Луценко Е.В. Системно-когнитивный анализ влияния довузовского дополнительного математического образования на успешность обучения в вузе / С. П. Грушевский, Е. В. Луценко, А. В. Назаров и др. ; под редакцией Е. В. Луценко; Министерство науки и высшего образования РФ, Кубанский государственный университет. — Краснодар: Кубанский гос. ун-т, 2023. — 131 с. — DOI: 10.13140/RG.2.2.28260.55684
11. Bogarín A. A Survey on Educational Process Mining / A. Bogarín, R. Cerezo, C. Romero // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. — 2018. — V. 8, № 1. — P. 1—17.
12. Predict students' dropout and academic success // kaggle. — URL: <https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention> (accessed 12.09.2023)
13. Интернет-репозиторий // Google Drive. — 2012. — URL: <https://drive.google.com/file/d/19C7-jALcQ3MIWGUefg4OqSDLE7AC2pU/view?usp=sharing> (дата обращения 12.09.2023)

### Список литературы на английском языке / References in English

1. Oficial'nyj sajt Nacional'nyj universitet «Vysshaya shkola ekonomiki» [Official website of the National University "Higher School of Economics"]. — URL: <https://www.hse.ru/> (accessed: 12.09.2023) [in Russian]
2. Realinho V. Predicting Student Dropout and Academic Success / V. Realinho, J. Machado, L. Bartista et al. // Data. — 2020. — V. 7. — P. 146. — DOI: 10.3390/data7110146
3. Zyateva O.A. Intellectual'nyj analiz dannyh pri kategorizacii prepodavatelej vuza na osnove naukometriceskih pokazatelej [Data Mining in the Categorization of University Teachers Based on Scientometric Indicators] / O.A. Zyateva, E.A. Pitukhin, I.V. Peshkova et al. // Inzhenernyj vestnik Dona [Engineering Bulletin of the Don]. — 2017. — No. 4. — URL: <https://ivdon.ru/magazine/archive/n4y2017/4580> (accessed: 12.09.2023) [in Russian]
4. Kotenko Yu.S. Problemy sovremennogo vuza i marketingovyje metody ih vyyavleniya i ocenki [Problems of a Modern University and Marketing Methods of Their Identification and Evaluation] / Yu.S. Kotenko, I.A. Nazvanova, M.G. Podoprigora // Inzhenernyj vestnik Dona [Engineering Bulletin of the Don]. — 2013. — No. 2. — URL: <https://ivdon.ru/ru/magazine/archive/n2y2013/1631> (accessed: 12.09.2023) [in Russian]
5. Yandybaeva N.V. Razrabotka programmnoho produkta dlya opredeleniya effektivnosti deyatel'nosti vysshego uchebnogo zavedeniya [Development of a Software Product for Determining the Effectiveness of Higher Education Institutions] / N.V. Yandybaeva, E.R. Kozhanova, V.A. Kushnikov // Vestnik Saratovskogo gosudarstvennogo tekhnicheskogo universiteta [Bulletin of the Saratov State Technical University]. — 2014. — V. 75, No. 2. — pp. 214-219. — URL: <https://elibrary.ru/item.asp?id=22567751> (accessed: 12.09.2023) [in Russian]
6. Yandybaeva N.V. Matematicheskie modeli, algoritmy i komplekсы программ для monitoringa effektivnosti obrazovatel'noj deyatel'nosti vuza [Mathematical Models, Algorithms and Software Complexes for Monitoring the Effectiveness of Educational Activities of the University] / N.V. Yandybaeva, V.A. Kushnirov // Problemy upravleniya [Problems of Management]. — 2015. — No. 1. — pp. 53-63. — URL: <https://elibrary.ru/item.asp?id=22958726> (accessed: 12.09.2023) [in Russian]
7. Yandybaeva N.V. Modelirovanie i prognozirovanie pokazatelej effektivnosti obrazovatel'noj deyatel'nosti vysshego uchebnogo zavedeniya [Modeling and Forecasting of Indicators of the Effectiveness of Educational Activities of Higher Educational Institutions] / N.V. Yandybaeva // Vestnik Mordovskogo universiteta [Bulletin of the Mordovian University]. — 2018. — V. 28, No. 1. — pp. 120-136. — DOI: 10.15507/0236-2910.028.201801.120-136. [in Russian]
8. Mayer R.V. Klassifikaciya uchebnyh faktorov metodom klasterного analiza [Classification of Educational Factors by the Method of Cluster Analysis] / R.V. Mayer // Problemy uchebnogo fizicheskogo eksperimenta: sb. науч. i metod. rabot [Problems of Educational Physical Experiment: collection of scientific. and the method. works]. — 1998. — Issue 5. — pp. 12-19. [in Russian]
9. Shevchenko V.A. Konceptiya postroeniya modeli priobreteniya znaniy studentami po discipline «Informatika» [The Concept of Building a Model of Knowledge Acquisition by Students in the Discipline "Informatics"] / V.A. Shevchenko // Vestnik HNADU: sb. науч. tr [Bulletin of KHNADU: coll. of scientific works]. — 2012. — Issue 56. — pp. 26-29. [in Russian]
10. Lutsenko E.V. Sistemno-kognitivnyj analiz vliyaniya dovuzovskogo dopolnitelnogo matematicheskogo obrazovaniya na uspešnost' obucheniya v vuze: monografiya [System-cognitive Analysis Effects of Pre-university Additional Mathematical

Education on the Success of Studying at the University] / S. P. Grushevsky, E. V. Lutsenko, A. V. Nazarov et al. ; edited by E. V. Lutsenko; Ministry of Science and Higher Education of the Russian Federation, Kuban State University. — Krasnodar: Kuban St. University, 2023. — 131 p. — DOI: 10.13140/RG.2.2.28260.55684 [in Russian]

11. Bogarín A. A Survey on Educational Process Mining / A. Bogarín, R. Cerezo, C. Romero // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. — 2018. — V. 8, № 1. — P. 1—17.

12. Predict students' dropout and academic success // kaggle. — URL: <https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention> (accessed 12.09.2023)

13. Internet-repozitorij [Online repository] // Google Drive. — 2012. — URL: <https://drive.google.com/file/d/19C7-jALc-Q3MIWGUefg4OqSDLE7AC2pU/view?usp=sharing> (accessed 12.09.2023) [in Russian]